PATHWAYS TO IMPROVING TRADITIONAL TRAVEL BEHAVIOR MODELS WITH TRAVEL-BASED MULTITASKING AND ATTITUDINAL DATA

A Dissertation
Presented to
The Academic Faculty

by

Aliaksandr Malokin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Civil and Environmental Engineering

Georgia Institute of Technology
May 2019

COPYRIGHT © 2019 BY ALIAKSANDR MALOKIN

PATHWAYS TO IMPROVING TRADITIONAL TRAVEL BEHAVIOR MODELS WITH TRAVEL-BASED MULTITASKING AND ATTITUDINAL DATA

Approved by:

Dr. Patricia L. Mokhtarian, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Patrick McCarthy
School of Economics
Georgia Institute of Technology

Dr. Giovanni Circella
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Kari E. Watkins
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Laurie A. Garrow
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: January 10, 2019

To entropy. To complexity.

ACKNOWLEDGEMENTS

This work is a product of outstanding, nurturing, and insightful environment of learning and collaboration that spanned over more than eight years across continents, states, and the multiverse of ideas. Even if the title page features only a single author, this manuscript is truly a work of many, directly and indirectly, to whom I am obliged and indebted forever.

First and foremost, I would like to thank my advisor, Dr. Patricia Mokhtarian, who guided me through my Master's and PhD degrees. With her patient mentorship and incredible southern charm, she introduced me to the way of thinking logically and statistically about the world– a gift that will remain with me long after the graduate school. Thank you for being kind to me and continue supporting me even while I myself thought it is not justified.

I am immensely grateful for the support Dr. Giovanni Circella provided me over these years. His gentle mentorship, endless optimism, and high degree of tolerance to my literary experiments helped me to become a better student.

I wish to thank my committee for providing valuable guidance and being very accommodating with the daring defense schedule I set up. It was my pleasure to work with Dr. Kari Watkins and to learn from her passion about everything bicycling and public transportation. I am thankful to Dr. Laurie Garrow for her incredible methodological insights and acumen for statistical programming. I am grateful to Dr. Patrick McCarthy for his refreshing econometric intuition and observations.

I would like to thank School of Civil and Environmental Engineering faculty, staff, and students for providing support, mentorship, and camaraderie. Specifically, I wish to express my gratitude to Dr. Ram Pendyala (currently at the Arizona State University), Marjorie Jorgenson, Robert Simon, and Dr. Mokhtarian's students: Atiyya Shaw, Ali Etezady, Sung Hoo Kim, and Xinyi Wang – guys, your own graduation is within your reach!

I have been happy to call Atlanta home for the last five years. Here, I have found good friends to share my love for cycling and board games with. Here, I have found my love, Nicole Johnston, to share my life with. Nikki, you have been a great inspiration to me!

Last but not least, I would like to thank my family and friends who have endured my lengthy trek to finish this manuscript. Mom and Dad, thank you for giving me an odds-defying start in life and supporting me through all these years half a world away. Anton, Dennis, and Vitaly, thank you for being awesome friends and shenanigans partners. Heather and Dan, thank you for always welcoming me in your home and for being my second family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AASHTO	American Association of State Highway and Transportation Officials
ACS	American Community Survey
ANL	Artificial Nested Logit
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ASC	Alternative-Specific Constant
AV	Autonomous Vehicle
CART	Categorical and Regression Tree
CCJPA	Capitol Corridor Joint Powers Authority
CIA	Conditional Independence Assumption
CV	Cross Validation
DA	Driving Alone
DAG	Directed Acyclic Graph
DOT	Department of Transportation
EM	Expectation Maximization
EV	Electric Vehicle
FHWA	Federal Highway Administration
GDP	Gross Domestic Product
GPS	Global Positioning System
HH	Household
HS	High School
HSR	High-Speed Rail

ICT	Information and Communication Technology
IIA	Independence of Irrelevant Alternatives
IT	Information Technology
ITS	Intelligent Transportation System
IV	Inclusive Value
IVTT	In-vehicle Travel Time
kNN	k-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
MCE	Misclassification Error
MNL	Multinomial Logit
MSE	Mean Squared Error
MSNCC	Multitasking Survey of Northern California Commuters
NHTS	National Household Travel Survey
NHTSA	National Highway Traffic Safety Administration
OVTT	Out-of-vehicle Travel Time
PC	Principal Component
PCA	Principal Component Analysis
RF	Random Forest
RHD	Random Hot Deck
RP	Revealed Preference
RV	Recreational Vehicle
SD	Standard Deviation
SOAV	Single Occupancy Autonomous Vehicle
SP	Stated Preference
SRAV	Shared-ride Autonomous Vehicle

SVM	Support Vector Machine
TOMNET	Center for Teaching Old Models New Tricks
TRB	Transportation Research Board
UC	University of California
VMT	Vehicle Miles Traveled
VO	Vehicle Ownership
VOTT	Value of Travel Time
VOTTS	Value of Travel Time Savings
WTP	Willingness to Pay
XGB	Extreme Gradient Boosting

SUMMARY

Transportation is undergoing extensive systemic changes: Information technologies are permeating through both transportation modes and people's activity patterns, as vehicle automation and ride-hailing/sharing platforms are "catching by surprise" our tested and proven planning and forecasting tools, while lifestyle preferences and behaviors of millennials (the largest generational cohort in the U.S.) are accelerating the digitalization of travel and may be redrawing land use patterns. In these tumultuous and uncertain times, ever more pressure is put on travel behavior models to understand and predict travel patterns, and to provide a foundation for sensible decision making.

Historically, regional transportation forecasting models used mostly socio-economic characteristics and relevant travel-related attributes to account for travel patterns. With the increased complexity and capacity for change of transportation systems, these factors could be insufficient for reliable policy and decision-making as the heterogeneity of travel preferences and experiences grows. Hence, the need for incorporating attitudinal data (an aggregate term for lifestyles, preferences, intentions, propensities, etc.), which underlies many travel-related decisions, into regional travel behavior models is especially strong now, and growing.

Accordingly, the main goal of the present dissertation is to contribute to the improvement of regional travel behavior models by investigating the influence of understudied behavioral drivers and increasing the availability of attitudinal data. This goal can be decomposed into two distinctive parts, among other ways unified through the use of a single attitudinally-rich dataset: (1) studying the effects of travel-based multitasking

on mode choice and the value of travel time (VOTT), and (2) developing an approach for porting attitudinal data from a small regional dataset to a large national sample.

For the first part of the objective, the empirical analysis is based on a survey of Northern California commuters (N > 2,000) that measures travel multitasking attitudes and behaviors, together with other attitudes, mode perceptions, and standard socioeconomic traits. We estimate a revealed preference mode choice model, which accounts for the impact of multitasking attitudes and behavior on the utility of various alternatives. Results show that the propensity to engage in productive activities on the commute, operationalized as propensity to use a laptop/tablet, significantly influences utility and accounts for a small but non-trivial portion of the current mode shares. For example, the model estimates that commuter rail, transit, and car/vanpool shares would respectively be 0.11, 0.23, and 1.18 percentage points lower, and the drive-alone share 1.49 percentage points higher, if the option to use a laptop or tablet while commuting were not available. Additionally, the work investigates the differences between millennials and older adults in the sample. Compared to non-millennials, the mode choice of millennials is found to be less affected by socio-economic characteristics and more strongly influenced by the activities performed while traveling.

For the second part of the objective, we transfer transportation-related attitudes from the same Northern California dataset to the 2009 National Household Travel Survey by augmenting both datasets with a large number of built-environment attributes and by applying machine-learning methods. Results indicate that the *pro-transit*, *pro-active transportation*, and *pro-density* attitudinal factor scores are predicted with the greatest precision; correlations of the predicted and observed scores are 0.564, 0.538, and 0.571, respectively. The performance of the transferred attitudes is measured by estimating linear regression models of vehicle ownership. The results show that in the source dataset the observed attitudes account for an 8.0% model lift (improvement in goodness of fit), while in the target dataset the predicted attitudes account for a 1.2–5.4% model lift.

The present study presents the valuable combination of a novel empirical application together with a data augmentation methodology that could be transferred to a variety of contexts. To our knowledge, it is the first study based on a revealed preference model to quantify the contribution of travel multitasking attitudes and propensities to mode choice. Also, it is the first empirical study to occupy the intersection of three timely travel behavior topics: the impact of activities while traveling on mode choice, the estimation of willingness to pay and VOTT, and the analysis of the travel behavior of millennials. Finally, this work acknowledges that with many transportation planning decisions requiring large-scale comprehensive datasets to be fed into travel behavior models, it would be difficult to achieve the introduction of a novel class of variables (e.g., activities while traveling) into the existing modeling pipelines. The proposed transfer learning framework targets this data unavailability and offers a way to synthesize promising variables into a practice-ready context.

CHAPTER 1. INTRODUCTION

1.1 Background and Motivation

In the field of transportation research, travel demand modeling is used to simulate complex, multi-actor transportation systems for forecasting, planning, engineering, and decision-making purposes. The quality and realism of the produced results and subsequent decisions depend on how well the underlying suite of models describes the transportation system. While functional forms of the models grounded on a comprehensive theoretical basis are important for accurate representation of the real world, one of the main factors behind a superior fit of the models is the relevance of data used in the model estimation process.

Historically, travel demand models used objective travel attributes (e.g., travel time, travel cost, route trajectories, and mode characteristics) and socio-economic attributes on household and personal levels as the primary explanatory variables. Over the years, these variables were demonstrated to be reliable predictors, albeit as proxies for the underlying behavioral drives and constraints that guide human decision-making. Being staple inputs in the models, the behavioral effects represented by these variables were studied thoroughly to become the foremost levers and targets for transportation planning, policy, and decision-making. At the same time, researchers continued scrutinizing alternative variables relevant to travel demand and behavior, some of them allowing capture of behavioral drives more directly and in finer detail.

Among those less conventional variables, behaviors that are logically related to travel, as well as attitudes (hereafter, we use this "umbrella term" to encompass people's beliefs, opinions, and preferences) have been long found to show a strong and unequivocal effect on manifested travel behavior. Although these unconventional variables have been defined and investigated across scores of smaller research datasets for decades, they are still virtually absent in large state- or nationwide samples, in effect making attitudes unavailable to the regional travel demand models used in practice. Both organizational inertia and budgetary constraints hinder the rapid "know-how" diffusion and adoption from a research lab to decision-making toolbox.

This research was partly motivated by a specific context in which the inclusion of such unconventional variables seems particularly desirable, namely, the influence of activities while traveling on travel behavior decisions – commute mode choice in the present study. The timeliness of this research is especially important as information technologies (IT) simultaneously alter both sides of the equation. On the one hand, the breadth and depth of activities that can be and are conducted while traveling (or travel-based multitasking) has exploded, with ever more portable and powerful computers, (smart-)phones, wearables, and other "gadgets" becoming virtually inseparable extensions of our identities (Han et al., 2017). At the same time, transportation modes themselves are not in stasis either, as IT optimizes existing alternatives (e.g., by providing mobility as a service – Zipcar, Uber, etc.) and creates new ones (e.g., by introducing autonomously driving vehicles – Waymo, Tesla, etc.). Digitalization of the transportation fleet has already become so advanced that the cybersecurity of intrinsically dangerous vehicles has become a pertinent issue (Petit and Shladover, 2015). Overall, one can argue that activities while traveling are making an increasingly large contribution to the utility of making a trip, even if the value of physically moving from point A to point B generally remains dominant.

Amidst these transformations, investigating the impacts of travel multitasking on travel behavior (i.e., commute mode choice in the present dissertation) could provide vital understanding of how activities while traveling could be leveraged in transportation forecasting and planning. Yet obtaining and using travel multitasking data effectively at the regional level would currently still require costly self-reported or observational data collection. Despite the development of internet-based surveys and smartphone-based lightning polls, a crucial problem with this type of data still exists: There is a direct relationship between the amount of useful information to be collected from respondents and their resource burden during this process, and correspondingly an inverse relationship between that burden and the likelihood of obtaining the desired information.

The status quo, however, could be ripe for dismantling. The deep digitalization and connectivity of the modern economy and people's everyday lives has presented opportunities not only to study how information and communications technologies (ICTs) affect travel behavior (e.g., influence of real-time wayfinding on traffic patterns, impacts of ride-sharing and ride-hailing platforms on mode choice and vehicle ownership, effects of various levels of autonomous driving technologies on safety and congestion, and – central to the present study – impacts of using smartphones and other ICTs on public transit ridership), but also to utilize them as a source of abundant, relatively inexpensive data. This data – for example, spatial positioning over time, accelerometer and compass readings, application usage, and social media activity – could have direct and indirect significance for modeling a travel behavior pattern of interest. However, whereas inferring a mode from trip trajectory, acceleration, and speed largely requires just a direct deduction from the user-generated data, measuring attitudes toward public transportation or propensity to work during travel could require more elaborate extraction mechanisms, which require establishing association patterns between raw data, attitudes and/or travel-related behaviors, and resulting travel behavior.

Accordingly, we can identify two complementing methods of improving data availability for transportation modeling: (1) at a practical scale, collecting new data that can be used to establish relationships between previously understudied (in a specific context) variables and travel behavior (i.e., *production*) and (2) bridging separate datasets to spread the "exclusive" data more widely (i.e., *transduction*). Executed jointly, the *Production-Transduction* framework would expand the data landscape by providing information for modeling and decision-making at a relatively low cost. Consequently, the *Production-Transduction* approach could be extended beyond transportation modeling data needs to other fields of knowledge where obtaining large amounts of human-specific data is prohibitively expensive, thus, contributing to our understanding and enhancing welfare.

A reasonable critic may say that the *Production* step is rather ordinary, and has been used ad infinitum since the advent of the Scientific Revolution, if not before. However, it is the specific context of the present research and its conjunction with the *Transduction* step that offer novelty. Conventionally, in the transportation field, after results of the *Production* step have been accepted by the community of theoreticians and practitioners, a wider study is commissioned to include the studied effects at the regional scale. Thus, the two steps rely on two distinct data collection efforts. This imposes temporal and financial constraints on the innovation of regional planning and forecasting tools. An alternative could include a knowledge transfer exercise in the form of capturing and passing on to regional models the conditional relationships that exist in the *Production* data. In other words, conditional relationships in the *Transduction* data (e.g., distribution moments, model parameters, etc.) are adjusted by the respective quantities observed in the *Production* data. One way to accomplish this is to transfer model parameters between the contexts, in which the *Production* data provide additional variables to account for parameter heterogeneity (Etezady et al., 2019). Another way is to transfer the variables that are responsible for heterogeneity directly by modeling their conditional relationship in the *Production* data and recreating it in the *Transduction* data, which is proposed by the current study.

This work develops and implements elements of both components of the *Production-Transduction* framework, to study its practicality and relevance for providing cost-effective ways to improve regional travel behavior models with attitudinal information. In the *Production* step the impact of travel-based multitasking propensities on mode choice is investigated to evaluate how the propensity to be productive during the commute influences the value of travel time, public transit ridership, and adoption of autonomous vehicles. The data for this step is available from the 2011 Multitasking Survey of Northern California Commuters (MSNCC, N > 2,000), which represents a "wide" design (over 1,000 original variables). Although in the present study this step does not directly feed into a regional model, the methodology (involving the creation and evaluation of alternative scenarios that are based on a revealed preference mode choice model containing attitudes) is broadly applicable and the specific, novel results are of direct interest to regional planning. Furthermore, as mentioned above, a variation on this step can be used to explicitly transfer to a regional model some updated parameters from a model estimated on a small sample that includes attitudes (Etezady et al., 2019).

The *Transduction* step develops and applies a knowledge transfer framework for enriching the NHTS 2009 dataset (N > 100,000, a "long" design), which (together with its predecessors and successor) is widely used for regional and national travel demand modeling, with selected attitudinal variables from the MSNCC. Although the variables used in the *Transduction* step are different from those whose influence is investigated in the *Production* step, the main contribution of the current study is developing a methodology that is robust with respect to a wide variety of inputs. To test it, the fidelity of the *Transduction* step is assessed by an external validation framework, which builds a vehicle ownership model using the NHTS data and relevant transferred variables.

1.2 Research Objectives

The main goal of the present dissertation is to contribute to the improvement of regional travel behavior models by investigating the influence of understudied behavioral drivers and by increasing the availability of attitude-based insights to those models. This goal could be decomposed into two distinctive parts, in accordance with the proposed *Production-Transduction* framework: (1) studying the effects of travel-based multitasking on mode choice and the value of travel time (*Production*), and (2) developing an approach for porting attitudinal data from a small regional dataset to a large national sample (*Transduction*). The main goal and sub-goals are achieved through a sequence of intermediary objectives:

Production sub-goals:

1. Obtain and prepare a comprehensive travel behavior dataset containing travel-based multitasking, attitudinal, socio-economic, and standard travel behavior data for modeling and transfer learning purposes. (Chapter 2)
2. Investigate how and to what extent the ability and propensity to travel-multitask influence the utility of various travel modes to an individual. (Chapter 2)
3. Analyze how and to what extent the travel behavior of Millennials (so called, "digital natives") is different from that of the older generations. (Chapter 3)
4. Examine how inclusion of travel-based multitasking and attitudinal variables influence derived values of travel time savings and willingness to pay measures in the mode choice model context. (Chapter 3)

Transduction sub-goals:

5. Develop a framework of transferring transportation-related attitudes from a smaller-scale travel behavior dataset to a large-scale travel survey. (Chapter 4)
6. Investigate to what extent the transferred variables improve travel behavior modeling in the context of a large-scale travel survey. (Chapter 4)

1.3 Dissertation Structure

The following three chapters (Chapters 2 through 4) of this dissertation are in journal format, i.e., each chapter is a self-contained journal article with the relevant inner structure. Accordingly, each chapter begins with an abstract, which is followed by the motivation, literature review, and methodological sections. Results and conclusions are, also, presented separately for each article in the respective chapter. However, the reference lists and appendices of the articles are combined and presented at the end of this document.

Chapter 5 provides the conclusions and directions for further research relevant to all three presented articles.

Chapter 2 and 3 comprise investigations associated with the research objectives that fall under the *Production* step, while Chapter 4 addresses the research objectives of the *Transduction* step.

In particular, in Chapter 2 we present a revealed preference mode choice model that accounts for the impact of multitasking attitudes and behavior on the utility of various alternatives. We then use the model to analyze several scenarios highlighting the potential near-term advantage of transit, and longer-term impacts of autonomous vehicles, associated with the ability to conduct activities while traveling. We find that the propensity to engage in productive activities on the commute, operationalized as using a laptop/tablet, significantly influences utility and accounts for a small but non-trivial portion of the current mode shares. The results empirically demonstrate the potential of a multitasking propensity to reduce the disutility of travel time.

In Chapter 3 we focus on the impact of travel multitasking on travel behavior and the value of travel time (VOTT) of the U.S. Millennial cohort. We estimate a revealed preference mode choice model and investigate the differences between millennials and older adults in the sample. Additionally, we conduct a sensitivity analysis to explore how incorporation of explanatory factors, such as attitudes and propensity to multitask while traveling, in mode choice models affects coefficient estimates and VOTT measures. Compared to non-millennials, the mode choice of millennials is found to be less affected by socio-economic characteristics and more strongly influenced by the activities performed while traveling. Young adults are found to have lower VOTT for both in-vehicle and out-of-vehicle travel time, even after controlling for demographic traits, personal attitudes, and the propensity to multitask.

In Chapter 4 we evaluate approaches to informing one dataset (the NHTS 2009) with knowledge (general transportation-related attitudes) from another (the MSNCC) and we evaluate the performance of the knowledge transferred into the NHTS dataset. Accordingly, the set of common variables is first augmented with a large number of built-environment attributes. Then, after applying machine-learning methods trained on the MSNCC data, we predict attitudinal variables in the NHTS dataset. The performance of the transferred attitudes is measured by estimating linear regression models of vehicle ownership. Although initial results are modest, we believe they show substantial promise, and the process has identified a number of opportunities for improvement and further research.

1.4 Major Contributions

The present dissertation possesses a valuable combination of novel research and applicable methodology that could be transferred to various contexts. This work makes three major contributions.

First, to our knowledge, this is the first study (Chapter 2; Malokin et al., 2019) based on a revealed preference model involving attitudes and mode perceptions to quantify the contribution of travel multitasking to mode choice.

Second, this is the first empirical study (Chapter 3; Malokin et al., 2018a) to occupy the intersection of three timely travel behavior topics: the impact of activities while traveling on mode choice, the estimation of willingness to pay and VOTT, and the analysis of the travel behavior of millennials. Although numerous studies have speculated that VOTT may be different for millennials and generations to come due to the ability to multitask while traveling, this is the first known empirical confirmation of that speculation. The strong impact on mode choice of the propensity to use a laptop while commuting found by this study prompts further research on the whole gamut of possible activities while traveling, as ICTs are integrating more seamlessly into our everyday lives and autonomous vehicles are starting to roam streets and highways all over the world.

Third, this work also realizes that with many transportation planning decisions requiring large-scale comprehensive data to be fed into travel behavior models, it would be difficult to achieve the introduction of a novel class of variables (e.g., activities while traveling) into the existing modeling pipelines. The proposed transfer learning framework (Chapter 4; Malokin et al., 2017b and Malokin et al., 2018b) targets this data unavailability and offers a way to synthesize promising variables into the practice-ready context.

The potential impact of the introduction of travel-based multitasking indicators and transportation-related attitudes into the travel demand modeling context via the approaches proposed in the current document could result in better simulation of travel behavior at micro and macro levels. Such models would be better equipped to analyze implications of the pressing issues in transportation: residential location choice, advent of autonomous driving, spread of transportation network companies (TNCs; e.g., Uber, Lyft, etc.), implementation of smart growth policies, and mode shifts towards active transportation alternatives (e.g., walking and bicycling), among others. The proposed approach provides a low cost, rapidly-developed pipeline between cutting-edge research and the real-world projects and policies implemented by transportation practitioners.

CHAPTER 2. IMPACT OF TRAVEL-BASED MULTITASKING ON MODE CHOICE

Malokin, Aliaksandr, Giovanni Circella and Patricia L. Mokhtarian (2019) How do activities conducted while commuting influence mode choice? Using revealed preference models to inform public transportation advantage and autonomous vehicle scenarios. *Transportation Research Part A* **124**, 82-114.

2.1 Abstract

From early studies of time allocation onward, it has been acknowledged that the "productive" nature of travel could affect its utility. Currently, at the margin an individual may choose transit over a shorter automobile trip, if thereby she is able to use the travel time more productively. On the other hand, recent advancements toward partly/fully automated vehicles are poised to revolutionize the perception and utilization of travel time in cars, and are further blurring the role of travel as a crisp transition between location-based activities. To quantify these effects, we created and administered a survey to measure travel multitasking attitudes and behaviors, together with general attitudes, mode-specific perceptions, and standard socioeconomic traits (N = 2229 Northern California commuters). In this paper, we present a revealed preference mode choice model that accounts for the impact of multitasking attitudes and behavior on the utility of various alternatives. We find that the propensity to engage in productive activities on the commute, operationalized as using a laptop/tablet, significantly influences utility and accounts for a small but non-trivial portion of the current mode shares. For example, the model estimates that commuter rail, transit, and car/vanpool shares would respectively be 0.11, 0.23, and 1.18 percentage points

lower, and the drive-alone share 1.49 percentage points higher, if the option to use a laptop or tablet while commuting were not available. Conversely, in a hypothetical autonomous vehicles scenario, where the car would allow a high level of engagement in productive activities, the drive-alone share would increase by 1.48 percentage points. The results empirically demonstrate the potential of a multitasking propensity to reduce the disutility of travel time. Further, the methodology can be generalized to account for other properties of autonomous vehicles, among other applications.

2.2 Introduction

Multitasking (doing multiple activities "at the same time") is a common feature of modern life, whether viewed as an annoying distraction, a means of increasing productivity or enjoyment, or both. There is a sizable and growing literature on multitasking in general (e.g., König and Waller, 2010), and in contexts such as work (e.g., Bluedorn and Martin, 2008; Chesley, 2014) or "media multitasking" (e.g., Wallis, 2010) in particular, but the study of activities conducted while traveling is a relatively young area of research (a comprehensive review of travel multitasking studies to date has been conducted by Keseru and Macharis, 2017). Multitasking has been thought to positively affect the (dis)utility of the trip (Mokhtarian and Salomon, 2001; Kenyon and Lyons, 2007; Wardman and Lyons, 2016) and thence the evaluation of travel time for a trip (a recent International Transport Forum Roundtable was devoted to the subject of "Zero Value of Time"; see https://www.itf-oecd.org/zero-value-time-roundtable, accessed November 19, 2018). At the margin, for example, some individuals may choose transit over the automobile for a given trip, even though the transit alternative takes longer, if in so doing they are able to use the travel time more productively.

These effects are expected to become even more relevant in future decades. One promise of partly- and fully-automated vehicles is to reduce the need for drivers to “pay attention to the road”. This, among other effects, will extend to private vehicles the hands-free advantage hitherto enjoyed by public transit, thus potentially allowing motorists to accrue the positive utility of travel-based multitasking (Anderson et al., 2014; Wagner et al., 2014). In this future, time slots that were previously almost exclusively occupied by travel will dissolve into more permeable channels permitting overlapping continuity of activities. In other words, travel will (often) lose its place as a primary activity of its own: activities that were previously possible only at the trip origin or destination (or could take place only when traveling as a passenger rather than a driver), such as relaxing or working with clients, could happen also aboard personal vehicles.

This study investigates the impacts of activities carried out while traveling (travel-based multitasking) on mode choice, specifically in the context of the daily commute. To do this, we created and administered a survey to measure multitasking attitudes and behavior while commuting, together with general attitudes, mode-specific perceptions, and standard socioeconomic traits (N = 2229 Northern California commuters). We used this dataset to estimate a revealed preference (RP) multinomial logit (MNL) mode choice model (Ben-Akiva and Lerman, 1985) that accounts for the impact of multitasking attitudes and behavior on the utility of various alternatives – to our knowledge, the first revealed-preference model to do so. We then used the model to analyze several scenarios highlighting the potential near-term advantage of transit, and longer-term impacts of autonomous vehicles, associated with the ability to conduct activities while traveling.

The paper addresses several research questions: after controlling for the conventionally included mode attributes and sociodemographic traits, as well as other (primarily attitudinal) variables expected to influence mode choice, how and to what extent do the ability and propensity to perform tasks while traveling influence the utility of various travel modes to an individual? What current share of public transit ridership could be attributed to travel-based multitasking? And what potential ridership could be captured by changes in the multitasking conduciveness of these modes? Finally, how would autonomous vehicles affect the mode split if their occupants could fully devote their attention to non-travel activities?

The remainder of the paper is organized as follows. Section 2.3 briefly reviews the literature, focusing specifically on the impact of travel-based multitasking on travel utility. Section 2.4 describes the empirical context of the study, including the data collection effort and the sample characteristics. We then present an overview of our methodological approach, and discuss the construction of mode-specific multitasking propensity measures, in Section 2.5. The mode choice model specification and the discussion of the effects of the explanatory variables are the subject of Section 2.6. In Section 2.7, we develop a set of transit-related and autonomous-vehicle-oriented scenarios, showcasing the potential shifts in mode shares attributable to multitasking factors. Finally, Section 2.8 presents some conclusions and future research directions. Appendix A provides additional technical details, including a discussion of issues associated with using a nested logit model instead of the sequential process we adopt, and proof that coefficients in a nested logit model differ when variables in (binary choice) lower nests are associated with different alternatives.

2.3 Literature Review

Within the past few years, growing attention has been paid to the impact of multitasking on travel behavior. For example, Guo et al. (2015) observed and surveyed 3425 students who used the college bus system in Vancouver, British Columbia with respect to their participation in passive/active, information and communication technology (ICT)-based/non-ICT-based, and smart-function/ non-smart function ("dumb phone") activities while riding and waiting. Even though the study focused on a very specific segment of the population (young adults who use public transportation and are often very familiar with ICT devices), the authors pointed to the importance of the temporal dimension (i.e., when and for how long activities are performed) in studying the effects of travel-based multitasking on the travel experience. The authors found that only 30% of the people who owned smartphones used them, despite large shares of riders engaging in non-passive activities while taking the bus (60%) or waiting for it (47%).

Tang et al. (2018) surveyed 901 passengers of high-speed rail (HSR) between Shanghai and Nanjing, and developed quadrivariate probit models of the engagement in four types of activities on the trip (ICT work, ICT non-work, paper work, and non-work paper reading). Nearly three-fifths of the sample engaged in ICT-based work (in contrast to European studies of conventional trains but in keeping with the higher-income clientele of HSR), and more than three-quarters engaged in ICT-based non-work activities. Interestingly, having a laptop along on the trip increased the propensity to conduct ICT-based work only for non-business travelers (signifying an intention to work while traveling for personal reasons), not for business travelers (many of whom may have brought the laptop primarily for activities at the destination).

The link between travel-based multitasking and the value of travel time was hinted at as early as 1965 (Becker), touched on by Mokhtarian and Salomon (2001), and elaborated conceptually by Lyons and Urry (2005). Watts and Urry (2008, p. 860) continued the discussion, arguing that travel time is by no means universally “wasted, dead, or empty”. Accordingly, a number of studies have empirically analyzed the link between multitasking and travel time. For example, Ohmori and Harata (2008) developed a descriptive analysis of activities conducted while traveling among 503 Japanese train riders and discussed the dependence of activity engagement level on travel time (e.g., using ICT for work purposes is more common during long trips). Using a scobit model estimated on a sample of 523 Japanese bus users, Zhang and Timmermans (2010) found that engaging in more activities while traveling decreases the sensitivity to changes in travel time. Thus, travel-based multitasking is likely to partly offset the negative impact of travel time on the utility of a travel mode. Studying ICT usage among Norwegian train commuters (N = 289) and business travelers (N = 245), Gripsrud and Hjorthol (2012) found that advanced planning and laptop usage increased the probability of getting work done (more so for business travelers) during the trip. Additionally, laptop usage was linked with a more positive subjective valuation of travel time for business travelers.

Several studies (Ettema et al., 2012; Susilo et al., 2012; Rasouli and Timmermans, 2014; Mokhtarian et al., 2015; Singleton, 2018) have explored the impact of activities conducted while traveling on the subjective evaluation of a trip experience. Among these, Rasouli and Timmermans (2014) found positive associations of working and shopping online, reading, and obtaining travel information, with improved perceptions of the trip experience among 98 Dutch participants in a three-month long GPS-based travel diary study. Similarly, in a study of 400 South Korean travelers, Rhee et al. (2013) included several activities conducted while traveling (e.g., talking on phone, chatting with passengers, and using an ICT device) in models of reported attitudes towards traveling, comparing automobile and public transit users. However, the reported activities were not universally available to users of both transportation modes, which inhibits an adequate side-by-side comparison. For example, while talking to other passengers had the same positive association with feeling happy for both automobile and transit users, social networking through ICT devices was investigated in the study (and was statistically significant and negative) only for automobile users.

At least two studies have estimated a monetary value of travel time savings (VOTTS) in the context of multitasking. Ettema and Verschuren (2007) relied on stated preferences (SP) and found that among 226 Dutch public transit riders, polychronic individuals (i.e., those with a more positive inclination toward multitasking) had lower VOTTS. Concluding, the authors warned about substantial exogenous heterogeneity of travelers' VOTTS. Other stated-preference studies also controlled for the activities conducted while traveling when assessing the demand for Wi-Fi on commuter rail (Connolly et al., 2009) and mode choice under urgent work tasks during the commute (van der Waerden et al., 2010). Varghese and Jana (2018), on the other hand, took a revealed preference approach to quantifying the VOTTS while accounting for multitasking. They segmented their sample of motorized trips in Mumbai, India into those on which multitasking occurred (N = 2037) and those where it did not (N = 913), and estimated separate mixed MNL mode choice models for each segment. They found that that the mean VOTTS was 26% lower for the trips on which multitasking occurred.

In a recent paper, Zheng et al. (2016) studied the utility of "laptop stations" on buses and urban trains in Australian cities. In the SP portion of the hybrid (SP/RP) survey, the authors asked more than 6,700 respondents whether the availability of a laptop station would influence their mode choice (bus/ train/ car). The results of the random effects logit model showed that the dummy variable (presence/ absence of a laptop station) was significant only in the train utility function: counterintuitively, the normally distributed random coefficient had a negative mean (apparently indicating that the presence of a laptop station *decreased* the utility of train), albeit a large standard deviation (meaning that the coefficient would be positive for a sizable fraction of the sample). Further, the utility of having a laptop station was found to be influenced by the trip purpose being commuting (positively) and by high income (negatively). The heterogeneity among respondents was additionally demonstrated by the willingness to pay for a train laptop station, which fluctuated near zero dollars and was negative for many survey-participants.

In sum, the burgeoning literature on travel-based multitasking exhibits considerable diversity with respect to geographic and cultural context, modes studied, activities examined, dependent variables analyzed, and whether stated or revealed preference formats were used; most studies do not segment the analysis on (or in some cases, even measure) trip purpose. All of this means that while prior research has certainly informed the present study, direct comparisons are problematic. While building upon the accumulating literature (including our own conceptualization of multitasking, in Circella et al. 2012), the current study is unique in its development of a revealed preference model of primary commute mode choice, incorporating perceptions of the "multitaskability" of each mode and estimated propensities to multitask on each mode, together with conventional measures of travel time and cost, to quantify the influence of travel multitasking on the disaggregate utility, and the resulting aggregate share, of each mode.

2.4 Empirical Context

To keep the scope of the study manageable, and in view of the importance of commuting as a daily anchor at the personal level and a key generator of congestion at the societal level, we chose to focus the transportation context of the study on commute trips. Among short-distance trips, although commuting nominally accounts for only about 15.61% of the personal transportation in the U.S. (U.S. Department of Transportation, Federal Highway Administration, 2009), many other trips are linked to the commute, it is typically the longest trip made on a frequent basis, and given its temporal peakedness, it is a major source of congestion and thence emissions. With respect to long-distance travel, although in many countries the train is a viable alternative to the car, that is not the case in the U.S. for the most part. The choice between air and car (when there even *is* such a choice) is made primarily on the basis of travel time and cost, with little room for desired travel multitasking to influence the decision. At the same time, in view of the latter considerations, most public transportation systems in the U.S. offer their best levels of service during commute peak hours. The frequency and length of the trip, together with the relative attractiveness of transit for such trips (compared to many other trip purposes), mean that the opportunities for productive travel multitasking, and the competitive appeal of transit for that reason, are generally highest for the commute trip (among short-distance trips).

Accordingly, our desired population consisted of commuters (including working college students) living in Northern California, with a particular but not exclusive focus on commuters traveling on the Sacramento – San Francisco Bay Area transportation corridor (the study area was chosen for geographic convenience, while the authors were affiliated with the University of California, Davis). Data collection was carried out in fall and winter of 2011-2012, using both paper and online versions of the survey (Neufeld and Mokhtarian, 2012). The survey was a single questionnaire that was administered once.

We used a variety of sampling approaches, including choice-based sampling (i.e., contacting people in the process of using their respective commute modes), mailing paper versions of the questionnaires to the addresses of a random sample of study area residents, and distributing links to the online surveys through employers'/ affiliated organizations' email lists and websites. Finally, we used the services of a commercial firm, Survey Analytics (https://www.surveyanalytics.com), to circulate the questionnaires to an appropriately filtered subsample of their paid panel members. Our goal was not to achieve a sample that was completely representative of the population of interest. Rather, we needed a sample with "enough" (a few hundred) users of each mode of interest to produce robust statistical results. In fact, the focus of the study lies in investigating the relationship of multitasking to mode choice, and the use of covariates in the estimation of the model – together with the weighting described in Section 2.6.1 – can largely control for biases due to the non-representativeness of the sample. Although any single sampling method would have been less-than-optimal if used in isolation, the combination of diverse methods helped alleviate the limitations of each, and has produced the desired diversity with respect to commute mode choice and other characteristics in the final sample.

Geographically, origins and destinations of the sampled commuters were unevenly distributed over several dozen Northern California counties. However, we were mainly interested in commutes within major agglomerations in the region. Therefore, only those respondents who commuted within 16 counties were included in the final sample for this study: the nine Metropolitan Transportation Commission counties (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano and Sonoma), the six Sacramento Area Council of Governments counties (El Dorado, Placer, Sacramento, Sutter, Yolo and Yuba), and San Joaquin County.

The final sample size for this study is 2229, after filtering out apparent mode captives and out-of-region, inconsistent, or frivolous respondents, as well as cases that were severely incomplete on key variables. Because of our sampling strategy, the sample descriptives (Table 2.1) can differ greatly from those of the general population. In particular, the sample considerably underrepresents drive-alone commuters, and overrepresents users of other modes.

Table 2.1 – Selected characteristics of the sample and population

Characteristic (sample size)	N (%)	Characteristic (sample/pop. size)	N (%)
Gender (2209)		**Commute total travel time distribution (2229)**	
Female	1370 (61.5)	Less than 15 mins	349 (15.7)
		15 to 30 mins	602 (27.0)
Age (2216)		31 to 45 mins	437 (19.6)
18 to 24	104 (4.7)	46 mins to 1 hour	284 (12.7)
25 to 40	750 (33.6)	1 to 1½ hours	323 (14.5)
41 to 64	1276 (57.2)	1½ to 2 hours	142 (6.4)
65 to 74	78 (3.5)	More than 2 hours	92 (4.1)
75 or older	8 (0.4)		
		Sample commute mode shares (2229)	
Education level (2229)		Biking	192 (8.6)
Some grade/ high school	3 (0.1)	Commuter rail	176 (7.9)
High school diploma	64 (2.9)	Transit[b]	649 (29.1)
Some college/ technical school	515 (23.1)	Shared ride[c]	355 (15.9)
4-year college degree	714 (32.0)	Driving alone	857 (38.4)
Some graduate school	241 (10.8)		
Complete graduate degree(s)	692 (31.0)	**Population commute mode shares (4,119,532)[a]**	
		Biking	63,187 (1.5)
Occupation (2221)		Commuter rail	29,508 (0.7)
Clerical/ administrative support	342 (15.3)	Transit	336,721 (8.2)
Homemaker	8 (0.4)	Shared ride	513,277 (12.5)
Manager/ administrator	375 (16.8)	Driving alone	3,176,839 (77.1)
Production/ construction	37 (1.7)		
Professional/ technical	1114 (50.0)	**Characteristic (sample size)**	**Sample mean**
Sales/ marketing	79 (3.5)		
Service/ repair	51 (2.3)	**Household size (2216)**	2.69
Student	189 (8.5)		
Other	26 (1.1)	**Number of operational household vehicles (2206)**	2.08
Annual household income (2142)			
Less than $25,000	127 (5.7)		
$25,000 to $49,999	313 (14.0)		
$50,000 to $74,999	436 (19.6)		
$75,000 to $99,999	414 (18.6)		
$100,000 to $124,999	358 (16.1)		
$125,000 or more	494 (22.2)		

[a] Population commute mode shares for the 16 Northern California counties of the study area were obtained from the Census Transportation Planning Products, available at http://ctpp.transportation.org/Pages/default.aspx, based on ACS 2006–2010 data.
[b] Includes local bus (current sample share 0.0557), express bus (0.0703) and light rail/subway (0.1783).
[c] Includes car/van driving with passengers (current sample share 0.0693), and carpool/vanpool/shuttle passenger (0.0849).

In addition to the socio-economic attributes, the collected data contains responses to various attitudinal statements, which were factor-analyzed to reveal the underlying attitudinal constructs (the constructs appearing in the final models are shown in Table 2.2 and Table 2.3). The factor analyses were performed on a cleaned dataset with a larger number of observations (all potentially eligible for future study, e.g., including non-working students; N ~ 2,800) by using principal axis and maximum likelihood methods for factor extraction, with oblique rotation and Bartlett factor score computation. With respect to the mode perceptions shown in Table 2.2, respondents were asked to rate multiple modes on parallel sets of attributes such as cost and comfort. For the factor analysis, instead of treating the parallel sets of responses as multiple variables for the same case (person), the variable sets were stacked "mode over mode", with each person-mode combination constituting a case. This was done so that the same factor structure would be obtained across modes (e.g., so that "comfort" would be associated with the same factor for all modes). One mode perception, namely its multitaskability, did not load onto any factor and was therefore included in the model as a stand-alone variable after being standardized for consistency with the factor scores.

Similarly, with respect to four of the time-use constructs shown in Table 2.3, parallel statements were presented for whether respondents felt they *must* engage in those behaviors, and whether they *wanted* to do so; these items were also stacked and factor-analyzed to have the same structure across those two variations on the question. The multitasking statements in Table 2.3 are those comprising the two main polychronicity scales (batteries of questions designed to measure a person's inclination to multitask) established in the literature (Bluedorn et al., 1999; Lindquist and Kaufman-Scarborough, 2007). Detailed reports on the factor analyses are available from the authors upon request.

Including the attitudinal variables enhances the estimated models in two ways: (1) it reduces the biases in the estimated coefficients of the *other* variables (notably, but not exclusively, the socio-demographic variables), which would otherwise be partially accounting for the explanatory power of the (missing) attitudes with which those other variables are correlated; and (2) it contributes substantial additional independent explanatory power to the model.

Objective mode attributes, specifically travel time and travel cost (averaged between morning and afternoon commutes), were obtained in post-processing, using fastest routes as suggested by Google Maps, and necessarily involving a number of assumptions as detailed below. Biking incurs a constant cost of $0 and travel time accounts for topography and accessible infrastructure, assuming an average speed on flat land of about 12 mph. Public transportation alternatives (commuter rail and transit) could be represented by a sequence of private and collective modes (along with walking and waiting episodes). Such alternatives are considered to be available for a commute if the aggregate travel time on collective modes and the associated wait time is over 50% of the total travel time (to exclude these modes from the choice set when lengthy access/egress times by private modes such as walking or car would be required), and if the reported work location can be reached by 9 am within a "reasonable" time (liberally set at 3 hours, in view of the geographic expanse of the commute shed for the region). Travel cost for the collective modes is determined by a summation of costs associated with the various modes involved in the trip. Cost minimization, such as a choice of the best pass (single ride, weekly, monthly) based on the reported commuting frequency and inter-agency ticket honoring, is applied. Travel time for driving alone is calculated as an average over several samples of real traffic conditions for AM and PM peaks. Travel cost for driving alone combines fuel (fuel efficiency is inferred via reported vehicle make and model), tolls, and parking costs. Shared ride travel time computation is similar to that for driving alone except for two details: 5 minutes were added to account for additional pick-up and drop-off times, and (where available) high-occupancy vehicle lanes were acknowledged through assuming free-flow speed over these segments. Further, shared-ride total travel cost, calculated similarly to that for driving alone, is divided by the average shared-ride occupancy for the region.

Table 2.2 – General attitudinal and mode perception constructs pertinent to the current study

Constructs	Statements[a]	Loadings[b]
	General attitudes[c]	
Pro-technology	I like to be among the first to own new electronic products.	0.755
	I like to track the development of technology.	0.747
	I often introduce new trends to my friends.	0.577
	The internet makes life more interesting.	0.343
	Technology brings at least as many problems as solutions.	–0.305
Pro-active modes	I like the idea of walking (or biking) as a means of transportation.	0.895
	I prefer to walk or bike rather than drive whenever possible.	0.767
	I like the idea of living in a neighborhood where I can walk to the grocery store.	0.420
Pro-transit	I prefer to take transit rather than drive whenever possible.	0.739
	I'd rather drive than travel by any other means.	–0.588
	I like the idea of driving as a means of travel for me.	–0.536
	I like the idea of transit as a means of travel for me.	0.510
Travel is wasted time	I generally enjoy the act of traveling itself.	–0.774
	The act of traveling is boring.	0.710
	Time spent traveling is generally wasted time.	0.592
	The only good thing about traveling is arriving at your destination.	0.567
	I sometimes travel more than I have to, because I want to.	–0.389
	To me, a car is mostly just a way to get from place to place.	0.308
	Mode perceptions[d]	
Mode convenience	Ability to run errands on the way to/from work	0.897
	Privacy	0.789
	Availability when needed/wanted	0.715
	Ability to carry things with me	0.591
	Door-to-door travel time	0.421
	Reliability	0.411
	Comfort	0.342
	Effect on the environment	–0.308
Mode benefit /cost	Effect on the environment	0.800
	Cost	0.626
	Avoiding congestion	0.583
	Amount of physical activity involved	0.557
	Ability to carry things with me	–0.311

Table 2.2 (continued)

Constructs	Statements[a]	Loadings[b]
Mode comfort	Safety	0.688
	Traveling in poor weather conditions	0.582
	Comfort	0.532
	Reliability	0.450
	Door-to-door travel time	0.376
	Ability to carry things with me	0.301
Mode multitask-ability	Ability to do things I need/want while traveling	stan-dar-dized single item

[a] A statement can load on more than one construct.
[b] Represents the degree of association between the statement and the construct. Only loadings greater than 0.3 are reported.
[c] Items measured on a 5-point Likert-type scale ranging from "Strongly disagree" to "Strongly agree".
[d] Items measured on a 5-point ordinal scale ranging from "Very bad" to "Very good". Thus, all items are positively oriented. Positive loadings for inherently negative items such as "cost", or ambiguous items such as "effect on the environment" or "amount of physical activity involved", should be interpreted as meaning, "viewing [this trait] favorably will contribute to a higher score on the associated factor".

Table 2.3 – Multitasking and time use constructs pertinent to the current study

Constructs	Statements[a]	Loadings[b]
	Multitasking preference[c]	
Polychronicity	I prefer to do one thing at a time.	–0.761
	I like to juggle two or more activities at the same time.	0.732
	Doing two or more activities at the same time is the most efficient way to use my time.	0.725
	I am comfortable doing more than one activity at the same time.	0.668
	I typically do two or more activities at the same time.	0.646
	When I work by myself, I usually work on one project at a time.	–0.608
	I believe it is best to complete one task before beginning another.	–0.603
	I would rather complete parts of several projects every day than complete an entire project.	0.566
	I believe people should try to do many things at once.	0.543
	I seldom like to work on more than a single task or assignment at the same time.	–0.538
	I believe people do their best work when they have many tasks to complete.	0.515
	I would rather complete an entire project every day than complete parts of several projects.	–0.492
	I believe it is best for people to be given several tasks and assignments to perform.	0.445
Multitasking is normative	I believe people do their best work when they have many tasks to complete.	0.800
	I believe people should try to do many things at once.	0.504
	I believe it is best for people to be given several tasks and assignments to perform.	0.433
	Time use	
Time spent working[d]	Amount of time you spend working	0.784
	Amount of time you spend relaxing	–0.452
	Amount of time you spend on the computer/phone/internet for work	0.415
Has to/ would like to work on commute[e]	Work during your commute	0.513
	Do "nothing" during your commute	–0.339
Has to/ would like to do recreation on commute[e]	Do recreational activities during your commute	0.641
	Socialize with other people while commuting	0.382
	Constantly be available to friends	0.299

Table 2.3 (continued)

Constructs	Statements[a]	Loadings[b]
Has to/ would like to multi-task at work[e]	On the job: work on several tasks in the time span of one day	1.022
	On the job: work on several tasks in the time span of one week	0.714
	On the job: work on several tasks in the time span of one hour	0.492
Has to/ would like to be available to people[e]	Constantly be available to friends	0.678
	Constantly be available to family	0.669
	Constantly be available to co-workers/clients	0.568

[a] A statement can load on more than one construct.
[b] Represents the degree of association between the statement and the construct. Only loadings greater than 0.3 are reported.
[c] Items measured on a 5-point Likert-type scale ranging from "Strongly disagree" to "Strongly agree".
[d] Items measured on a 5-point ordinal scale ranging from "Way too little" to "Way too much".
[e] Items measured on a 3-point ordinal scale ranging from "Generally no" to "Generally yes".

2.5 Methodological Approach

With the empirical context and set of available variables in mind, this section describes our methodology in greater detail. Section 2.5.1 presents the entire process, while Section 2.5.2 focuses on an important and novel component of the process: the measurement (estimation) of mode-specific propensities to engage in various activities on the commute. Taken together, this section offers a "blueprint" that could be replicated in numerous contexts both similar and dissimilar to the one of this study.

2.5.1 Overview of the Methodology

Narrowly construed, this study offers a methodology for assessing the implications for mode choice of the emergence of new technologies for travel-based multitasking, using a *revealed-preference* discrete choice model. Once the model parameters are estimated, carefully constructed scenarios allow for the evaluation of a "counterfactual present" (i.e., what mode shares *would have been, without the new technology*), as well as multiple "hypothetical futures" (i.e., what shares *could be, if currently familiar modes become more conducive to the use of the new technology*), all else equal. These applications obviously involve a number of assumptions on how the counterfactual present and hypothetical future *differ* from today's reality. But because they originate in a representation of actual present-day behavior, pivoting on what the counterfactual present and hypothetical future have *in common* with today's reality (including the present-day multitaskability of transit and shared-ride modes), we believe they offer a degree of verisimilitude not necessarily present in stated-response models that do not have that degree of commonality, and which require respondents to imagine a rather different world than today's. Nevertheless, the applicability of this approach will obviously be limited to the extent that future modes evolve into forms relatively *unfamiliar* today.

Furthermore, as applied in this study, the methodology considers primarily a single aspect that is not conventionally considered in a mode choice model: each mode's multitaskability. Even considering just the context of autonomous vehicles (let alone other transportation alternatives of the future), it is clear that many other aspects of a mode could be important to its adoption: safety perceptions, a "coolness" factor, congruity with self-identity, perceived desirability of sequential or simultaneous sharing of vehicles for that mode, and so on. In principle, the methodology can readily be expanded to incorporate any number of pertinent attitudinal constructs into the model. Survey design considerations, however, will likely limit the number that can practically be included.

Viewed even more broadly, however, a key element of the methodology is its approach to turning observations on *consequences of the chosen mode* (activities conducted on a specific commute, in this case) into *propensities for those consequences to occur if any particular mode were to be chosen*, and then incorporating those propensities into the mode choice model as explanatory variables (as explained in detail in Section 2.5.2). In effect, it is a way to capture the influence, on the choice to be made, of the *anticipated consequences* of each of the *possible* alternatives, using only the *observed consequences* of the *actual* choices. This approach could have applications in any number of contexts. For example, suppose we want to model the choice between store and online for a recent shopping activity. We could expect the likelihood of needing to return the item to affect the choice of shopping mode (or "channel", in marketing research parlance), but what we observe is *whether the item obtained via the chosen channel needed to be returned*. The methodology of this study offers a way to estimate the *likelihood of return for each shopping channel, if it were to be chosen*.

With these observations in mind, then, below we briefly recapitulate the main steps of the methodology. Note, again, that various simplifications, assumptions, and decisions will need to be made at each step. In the present study, for example, we focused on commute trips only, and only the primary commute mode (see Sections 2.4, 2.5.2, and 2.6 for a number of these assumptions etc.). Although such assumptions constitute limitations of the approach, they are consistent with the general character of models as being useful simplifications of reality.

1. Carefully consider the variables expected to influence mode choice in the study context: objective mode attributes, mode-specific perceptions, other attitudes, and sociodemographic variables. Also consider the important *consequences* of mode choice, the *anticipation of which* might influence choice.

2. Design a survey to measure the variables identified in Step 1 as being pertinent to the choice process of interest, and administer the survey with attention to obtaining an ample number of cases choosing each mode.
3. As needed and available, supplement the self-reported information from the survey with external data, as we did to obtain travel times and costs for all modes (available to an individual) rather than only for the chosen mode.
4. As needed and appropriate, synthesize responses from individual attitudinal statements into composite, continuous-valued scores on attitudinal factors.
5. Develop mode-specific propensities to experience a given consequence ("use a laptop during the commute", in the present study; see Section 2.5.2 for details):
    a. Using only the choosers of mode *j*, estimate a model for whether the given consequence is experienced or not, as a function of explanatory variables available for everyone in the sample (both choosers and non-choosers of mode *j*).
    b. Using the model estimated for mode *j*, compute predicted probabilities of experiencing the consequence of interest while using mode *j*, for *all* cases who have mode *j* in their choice set (including both choosers and non-choosers of mode *j*).
    c. Repeat for each mode. The conclusion of this step will find each mode in a person's choice set to have an estimated "propensity to experience the consequence of interest" associated with it.
6. Develop a mode choice model, including as explanatory variables objective mode attributes, mode perceptions, other attitudes, sociodemographic variables, and the estimated propensity to experience the consequence of interest (Section 2.6). If choice-based sampling were used in Step 2, the sample should be weighted to properly replicate population mode shares.

7. Compute value of travel time savings and willingness-to-pay measures as desired (Section 2.6; Malokin et al, 2017a; Malokin et al, 2018a).

By manipulating the values of selected explanatory variables and/or coefficients, construct various scenarios representing plausible or instructive “hypothetical future” or “alternative present” cases. For a given scenario, use the model estimated in Step 6 (together with the manipulated inputs/parameters) to compute disaggregate predicted probabilities of choosing each mode. Aggregate those probabilities across the (weighted) sample to obtain mode shares associated with the scenario, and compare them to the benchmark shares (Section 2.7).

2.5.2 Estimating Mode-specific Propensities to be Engaged in Certain (Types of) Activities while Commuting

Multiple measures of multitasking are available in the data. First, respondents' personal orientation toward multitasking in general, i.e., their polychronicity, was measured as described in Section 2.4. Second, respondents were asked to rate each alternative mode on how well it offered the “ability to do things I need/want while traveling” (*mode-specific perceptions*, Table 2.2). Third, they indicated which of a number of different activities they performed on a single recent commute (*chosen-mode-based behavior*, Table 2.4). The mode-specific perception can be included in a model as either generic (with a constant coefficient across modes) or alternative-specific variables. The activities performed on a recent commute, however, are known only *after* the mode choice being modeled has been made, which makes them endogenous and therefore not directly suitable as explanatory variables *influencing* choice. Put another way: just as we need to know travel time and cost not only on the chosen mode but also on the alternative modes, so it is not enough solely to know what a commuter *did* on a *particular* mode; we also need to know *what she could have done* on *other* modes to know how travel-based multitasking would influence her mode choice.

There are multiple ways to incorporate the effects of mode-based multitasking behavior into a mode choice model. One conceptually elegant way is to view both decisions (which mode, and whether or not to use a laptop) as a multidimensional joint choice problem (Ben-Akiva and Lerman, 1985). Accordingly, one can specify a nested logit model where lower nests would represent the choice between using laptop or not for a given mode, and the upper nest represents the choice of mode. In this way, all parameters of the model are estimated simultaneously, and the use of full information for both choices yields efficient estimators. The inclusive value for each lower nest, representing (loosely speaking) the maximum expected utility of that nest, is fed into the utility function of the associated mode in the upper nest, so that the probability of choosing a given mode is influenced by the benefit the traveler expects to receive from the decision to multitask or not on that mode. The appendix to the paper (Section A.1) explains the issues associated with this approach, which are why we ultimately chose the alternative approach presented here.

Table 2.4 – Activities performed during commute

Activity (sample size)	Biking	Commuter rail	Transit	Shared ride	Driving alone
	Number of engaged commuters (% of mode choosers)				
Number of choosers:	(186-192)	(168-176)	(625-647)	(343-354)	(836-855)
Technological					
Smartphone[a] (2200)	20 (10.6)	122 (70.1)	297 (46.3)	132 (37.8)	241 (28.5)
Internet[a] (2205)	6 (3.2)	86 (49.7)	277 (42.9)	100 (28.6)	94 (11.1)
Reading electronically[a] (2181)	2 (1.1)	90 (52.9)	216 (34.1)	77 (22.4)	54 (6.4)
Gaming electronically[a] (2191)	2 (1.1)	42 (24.9)	147 (23.0)	39 (11.2)	24 (2.8)
Messaging[a] (2206)	14 (7.4)	127 (73.0)	334 (51.7)	140 (40.1)	158 (18.6)
Recreational					
Watching scenery/ people (2216)	154 (80.6)	134 (76.1)	479 (74.0)	223 (63.4)	377 (44.4)
Daydreaming (2208)	146 (76.4)	89 (51.7)	387 (59.8)	169 (48.4)	355 (41.8)
Exercising (2207)	185 (96.4)	14 (8.2)	47 (7.3)	5 (1.4)	13 (1.5)
Productive					
Writing electronically[a] (2179)	1 (0.5)	75 (43.6)	65 (10.3)	48 (14.0)	19 (2.2)
Laptop/ tablet[a] (2199)	1 (0.5)	82 (47.4)	61 (9.5)	65 (18.6)	31 (3.7)
Thinking/ planning[a] (2219)	159 (83.7)	135 (77.1)	475 (73.5)	267 (75.4)	651 (76.2)
Traditional					
Reading from paper[a] (2194)	2 (1.1)	113 (66.1)	347 (53.9)	48 (13.8)	38 (4.5)
Sleeping/ resting (2208)	1 (0.5)	100 (58.1)	252 (39.0)	95 (27.1)	16 (1.9)
Talking to strangers[a] (2198)	5 (2.6)	71 (41.3)	168 (26.3)	40 (11.5)	27 (3.2)
Writing on paper[a] (2181)	4 (2.1)	68 (39.5)	99 (15.5)	26 (7.5)	20 (2.4)
Talking to friends[a] (2201)	12 (6.4)	99 (57.2)	277 (43.3)	292 (83.2)	70 (8.2)
Gaming non-electronically[a] (2192)	2 (1.1)	9 (5.3)	24 (3.8)	17 (4.9)	14 (1.6)
Maintenance					
Eating/ drinking (2222)	19 (9.9)	122 (69.3)	100 (15.5)	145 (41.0)	424 (49.6)
Audio[a] (2218)	51 (26.8)	101 (51.7)	275 (42.7)	258 (72.9)	813 (95.1)
Grooming (2196)	0 (0.0)	23 (13.7)	35 (5.5)	23 (6.5)	56 (6.6)
Talking on phone[a] (2201)	22 (11.6)	99 (57.6)	213 (33.4)	84 (23.9)	281 (33.1)
Navigating[a] (2162)	5 (2.7)	18 (10.5)	45 (7.2)	46 (13.4)	118 (14.1)
Watching video[a] (2198)	2 (1.1)	37 (21.5)	59 (9.2)	20 (5.7)	27 (3.2)

[a] Originally, the activity had been reported separately for two purposes: work and leisure/ personal. For this analysis the purposes were combined.

This alternative approach is conceptually similar to the nested logit formulation, in that it allows the prospective benefit of multitasking on a given mode to feed into the utility function for that mode. It differs in that (a) we use a two-stage approach, analogous to the sequential (limited-information) rather than simultaneous (full-information) estimation of nested logit, and (b) the prospective benefit of multitasking on a given mode is reflected by the predicted probability (or *propensity*) of multitasking if that mode were to be chosen, rather than by the inclusive value function. In the latter respect, it is loosely inspired by (though not identical to) the propensity score regression approach to treating endogeneity bias in the context of treatment evaluations[1] (see, e.g., Newgard et al., 2004).

---

[1] In treatment evaluation studies, the target variable is an outcome of some kind (such as blood pressure, for medical applications), which is often modeled as a function of the treatment indicator (yes or no), plus a number of pertinent covariates. However, if treatment is not assigned randomly, then characteristics that differ between treated and untreated cases could confound the estimated effects of treatment. One remedy is to estimate a separate model of the probability of being treated or not, and then include that estimated probability of (or propensity for) treatment as another control variable in the outcome model. Doing so means that the remaining coefficients (in particular, that of the treatment indicator) can properly capture the effects of the associated variables (particularly, receipt of treatment or not) for people with the same propensity to be treated.

In our context, the outcome of interest is mode choice. The "treatment" loosely corresponds to "uses a laptop or not" – but only loosely. In our case the "treatment" clearly occurs after the outcome, which means that it is not causally consistent to have the treatment indicator be a predictor in the outcome model (as would be the case in a conventional treatment evaluation context). In a discrete-outcome context such as ours, the likelihood of "treatment" is also conditional on a specific discrete outcome occurring, which means that the treatment propensity model can only be estimated conditional on a specific outcome occurring (also unlike the conventional situation). Nevertheless (to the extent that the propensity model estimated for the choosers of mode j can be considered transferable to the non-choosers of mode j), we can treat the estimated propensity to use laptop on a given mode as a (counterfactual, for non-choosers of that mode) measure of the benefit the person would receive *if using that mode*, and include it as a statistical control in the utility function for the associated mode (so that the utility, in turn, represents "what the utility would be if that mode were to be used").

Is it reasonable to assume that the propensity model estimated for the choosers of mode j is transferable to the non-choosers of mode j? Taste heterogeneity between a mode's choosers and non-choosers is quite possible even with an ordinary mode choice model, but a model whose utility function coefficients differ between *choosers* and *non-choosers* of a given alternative (as distinct from one whose coefficients simply differ by alternative, *regardless* of whether that alternative is chosen or not) would not be estimable. Accordingly, *all* mode choice models assume that although the values of *explanatory variables* may differ between choosers and non-choosers of a given mode, the values of the *coefficients* of those variables do not. In the same way, our laptop usage models allow the resulting estimated propensities to differ between choosers and non-choosers (by virtue of the values of the models' explanatory variables differing), while assuming that the coefficients of the models used to estimate the propensities do not.

To implement this approach, we estimated the propensity to conduct a particular activity associated with a particular primary commute mode[2], as follows. For each mode-activity combination, we formulated a binary logit model using travelers' mode-specific involvement in each activity (=1 if reported, =0 otherwise), as the dependent variable. Individual characteristics such as socio-economic attributes, multitasking preferences, general attitudes and personality traits, time use expectations and preferences, and attitudes toward waiting were used as observed explanatory variables; although none of these variables differs by mode, their influence on utility (i.e., their coefficients in the models) can. The error term captures the net effect of all unobserved variables on the utility of performing the given activity when using the given mode; those unobserved variables include the intrinsic conduciveness of the mode to performing that activity. The model was calibrated on respondents who chose that mode, and the result was applied to predict the probability of performing that activity *if that mode were to be chosen* for all respondents, regardless of their actual mode choice (for additional information, see Berliner et al., 2015).

The two-stage approach we use here makes the assumption that the model of laptop usage on mode *j* that is estimated on choosers of *j* applies equally well to non-choosers of *j* (i.e. that the estimated *coefficients* are the same for both groups, although we stress that the distributions of the associated *explanatory variables* are allowed to, and almost certainly will, differ by group). The approach also results in a loss of efficiency for the

[2] We asked about activities conducted while commuting, but to keep the survey burden manageable, for multimodal commutes we do not know the specific mode in use at the time of the activity. Our mode choice model pertains to the *primary* commute mode, defined to the respondents as the mode that was used for the longest portion of the commute trip. Thus, we effectively assume that an activity of interest is being conducted on the primary commute mode, which may be incorrect for some activities. For example, the three respondents reporting using a laptop/tablet, but whose primary mode is "biking", may have used a tablet on a walk or transit passenger leg of the trip. On the other hand, they could also have been listening to music through earbuds attached to the tablet in their backpack as they cycled to work.

mode choice model estimation (because only information from the choosers of a given mode is used to estimate the propensity to multitask on that mode), and therefore the statistical tests of significance for those parameters should be considered approximate indications. Since for most of our results, however, statistical significance is far stronger than the typical 0.05 threshold, we believe that the essential nature of the estimated mode choice model is sound. Additionally, the two-stage approach allows us to implement weighting only for the mode choice model (see the related discussion in Appendix A), thus more appropriately representing many effects that influence laptop use on the smaller-share collective modes.

Among 23 reported activities, we selected the propensity to use a laptop, netbook, or tablet computer for work or personal purposes ("use a laptop", hereafter) for inclusion in the mode choice model specification. This decision was based on several reasons. For one thing, conceptually, personal computer usage could be strongly associated with a plethora of productive tasks that commuters would like to undertake to make more valuable use of their travel time (objectively and subjectively). This assumption is corroborated by the data: 61.5% of the respondents who used laptop reported "allows me to get more work done" to be among the benefits of the activities they do while commuting. For another thing, an exploratory factor analysis (Malokin et al., 2015) that we developed on the propensities to engage in activities while traveling showed close association between using a laptop and "writing/editing electronic documents": together with "thinking/ planning" and "reading electronic documents", they all load on one factor, i.e., "productive [activities]". While writing/editing electronic documents could be enabled by a (continuously increasing) variety of technological devices, usually a laptop computer (or a tablet) represents a major gateway for being productive, especially while traveling. From a general perspective, we can view the propensity to use a laptop/ tablet/ netbook during a trip as a *proxy* for the propensity to be productive while commuting on a given mode. More literally, however, we can view it as *only one of various ways* to be productive while commuting, and as such we can expect our results to understate the influence on mode choice of a propensity for productive travel multitasking. Either way, the laptop is merely one current medium of achieving such productivity, which will doubtless be at least partly supplanted by other media over time. To the extent that such new media will also allow for more and better ways to productively travel multitask, our results will be further understated. However, the *methodology* described in this study is robust with respect to advances in technology, and can readily be applied to new and improved media as they emerge.

Although the propensity to use a smartphone, another ICT-based activity which is even more commonly conducted while traveling than using a laptop (31.5% and 6.5% reported using a smartphone and laptop, respectively), was also tested in the mode choice model, it consistently produced coefficients with a negative sign, implying that using a smartphone *decreased* the utility of the given mode. Not only is this counterintuitive, but also, from the conceptual perspective, it seems unlikely that the decision to use a smartphone takes precedence over (and influences) the choice of a commute mode. Rather, it seems more plausible that the opposite direction of causality is indicated, meaning that commuters are more inclined to use smartphones on "lower-utility" modes, to help compensate for the greater disutility of those modes[3]. This finding and interpretation is consistent with others in the literature (e.g., Ettema et al., 2012; Mokhtarian et al., 2015). Consequently, we chose to exclude the smartphone propensity variable from the model. Keeping in mind that the data were collected in 2011, it is likely that smartphones have become much more prevalent as a productivity tool now than they were at the time (an example of the new media referred to above), so that results obtained with more recent data (potentially including the measurement of additional or alternative attitudinal constructs) may be different. However, it is also possible that during the commute, smartphones are still more often used for entertainment (games, texting, web-browsing) than for productivity – a useful subject for further research.

Model estimation results of the propensity to use a laptop while traveling on each mode are presented in Table 2.5[4] (an alternative way of presenting the results of the estimation – as utility function equations –can be found in the appendix, Section A.3) For economy of presentation we do not interpret the models here, but a full discussion/interpretation of these and other activity propensity models is found in Berliner

---

[3] An alternative argument could be that the result is merely a consequence of the coincidence that the modes with lower market shares – i.e., lower average utility – are also those more conducive to using a smartphone, but the same argument is true for using the laptop, which does not explain why the coefficients of the two variables have opposite signs.

[4] One point that may deserve mention here, however, is the use of commute distance rather than duration in the model. As a reviewer pointed out, duration is arguably the more relevant of the two measures. However, (1) distance and duration *are* highly correlated (above 0.93 for all modes), and (2) using distance allows us to create laptop propensities even for modes that are not currently in the individual's choice set (for example, in cases where there is no bus/light rail service that connects residence and workplace locations). Although such modes, by definition, will not come into play in the mode choice model, we considered it useful for our methodology to be able to create the laptop propensity in case those alternatives were to become available to the commuter in the future. For this reason, we chose to use distance as a very good (and always available) proxy for the conceptually more apt (but not always available) duration.

The same reason (2) applies to the decision of not using *mode multitaskability* perception in the binary logit model specifications, as we sacrifice an explanatory variable for the benefit of having a full set of propensity to use a laptop/ tablet measures.

et al., 2015. All mode-specific final specifications contained exogenous explanatory variables except for the biking model, which, due to the few cyclists reporting using a laptop during their commute, has only a constant term (yielding constant predicted propensities equal to the (low) share of bicycling commuters who use a laptop; see footnote 2). The goodness-of-fit measures, ranging from 38% to 84% of information explained (Hauser, 1978), are high in part because of the unbalanced shares of laptop choosers and non-choosers for most of the modes. Temporarily removing the constants (Mokhtarian, 2016) shows that the explanatory variables account for more than 97% of the explanatory power of the full model for the commuter rail and drive-alone modes, and as low as about 8% for the transit mode. In general, the models exhibit respectable predictive ability, considering their parsimonious nature and the inability of the available variables to fully capture the many factors behind specific multitasking behaviors.

In particular, it is worth pointing out that data were not available on crowdedness, ride bumpiness, and other travel experience variables, which could certainly be expected to influence the propensity to multitask in general, and to use a laptop in particular. However, the net influence (on using laptop) of the level of service and trip conditions experienced by the users of each mode is reflected in the constant terms of the mode-specific models of Table 2.5. Not surprisingly, the constants (which, of course, also include the effects of other unobserved variables) are negative (reflecting a lower propensity to use laptop) for all alternatives other than commuter rail.

Table 2.5 – Binary logit models of the mode-specific propensity to use a laptop, netbook, or tablet computer

Variables	Biking	Commuter rail	Transit	Shared ride	Driving alone
	General attitudes[a]				
Pro-technology	–[b]	–	0.549*** (0.120)	–	–
Travel is wasted time	–	–	–	0.564*** (0.168)	–
	Multitasking preference				
Multitasking preference (polychronicity)	–	–	0.241** (0.120)	–	–
Multitasking is normative	–	–	–	–	0.401** (0.184)
	Time use				
Time spent working	–	–	–	–	–0.372** (0.185)
Has to work on commute	–	1.148*** (0.209)	0.368*** (0.114)	1.262*** (0.189)	0.770*** (0.172)
Has to do recreation on commute	–	–	–	–	0.946*** (0.234)
Would like to do recreation on commute	–	–	–	0.685*** (0.225)	–0.389[c] (0.230)
Has to multitask at work	–	–	–	–0.456** (0.197)	–
Would like to be available to people	–	–	–	0.486*** (0.184)	–
Would like to take same route[e]	–	–0.543*** (0.203)	–	–0.383** (0.188)	
	Socioeconomic characteristics				
Female	–	–1.360*** (0.431)	–	–	–
Age	–	–0.049*** (0.015)	–	–	–
Hourly waged (=1 if 'yes',=0 otherwise)	–	–3.276** (1.265)	–	–	–
Vehicle age	–	–	–	–	–0.102** (0.041)
Annual household per capita income, $000	–	–	–	–0.021*** (0.006)	–
Travel distance, mi	–	0.026*** (0.007)	–	0.029*** (0.008)	–
	Constants				
Constant	–4.470*** (0.000)	0.313 (0.827)	–2.268*** (0.135)	–4.408*** (0.483)	–2.178*** (0.415)

Table 2.5 (continued)

	Biking	Commuter rail	Transit	Shared ride	Driving alone
N (for whom given mode is primary)	265	197	811	389	1001
Choosers (of laptop during commute)	3	95	95	72	37
$\mathcal{L}\,(\mathbf{0})$	–183.684	–136.550	–562.142	–269.634	–693.840
$\mathcal{L}\,(\boldsymbol{c})$	–16.426	–136.426	–292.922	–186.341	–158.328
$\mathcal{L}\,(\widehat{\boldsymbol{\beta}})$ w/o constants	–183.684	–84.192	–539.375	–191.799	–148.371
$\mathcal{L}\,(\boldsymbol{\beta})$	–16.426	–84.128	–272.025	–113.711	–132.445
ρ^2 ($\mathcal{L}\,(\mathbf{0})$ base) w/ ASC	0.9106	0.3839	0.5161	0.5783	0.8091
Adjusted ρ^2 ($\mathcal{L}\,(\mathbf{0})$ base)	0.9051	0.3326	0.5090	0.5449	0.7990
Share of explanatory power due to true variables[d], %	0.00	99.88	7.85	49.92	97.16

***, ** = significant at 1%, 5%.

[a] Effects of the variables are represented by an estimated coefficient and standard error (in parentheses).

[b] Dashes indicate coefficients that were constrained to be zero after they were found to have significance > 0.05.

[c] This coefficient is significant at the <0.09 level. It was more significant in preliminary specifications; however, after additional data cleaning to replace missing values, which increased the sample size, the coefficient exceeded the 0.05 threshold. It was maintained in the current specification for its conceptual merit.

[d] Defined as the ratio between the ρ^2 w/o ASC and the ρ^2 w/ ASC.

[e] Standardized response to this single item.

The estimated mode-specific propensities given the chosen mode are shown in Figure 2.1. The results are indeed interesting, showing, for instance, that the distributions of the laptop propensities for the rail and shared ride modes differ substantially between rail *choosers* and choosers of the other modes (as a result of differing values on the explanatory variables between rail choosers and others). In these two instances, the distribution for rail choosers has a uniform-like shape while for the remaining cases the distribution has a distinctive exponential decay or gamma-like shape. This signifies that a far greater share of rail choosers has a relatively high propensity to use laptops on the two modes that arguably permit it best (given that crowded conditions, shorter trips, and more frequent transfers on local transit are often not conducive to laptop use), compared to choosers of other modes. In other words, those who are most inclined to use laptops while commuting have been able, to a certain extent, to sort themselves into a mode that allows them to do so, whereas even if choosers of other modes happened to find themselves on commuter rail or sharing a ride, they would still not be highly inclined to use a laptop.

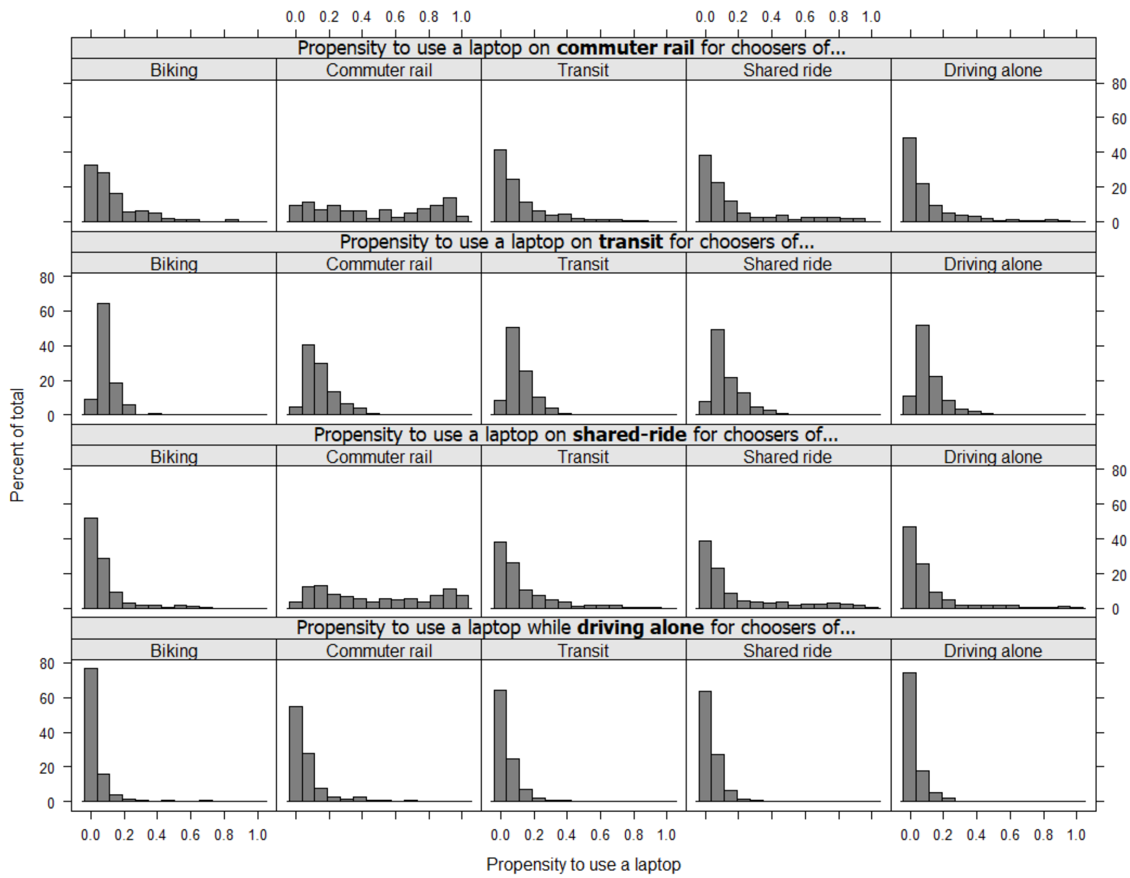

Figure 2.1 – The estimated mode-specific propensities to use a laptop (row) given the chosen mode (column)

2.6 MNL Mode Choice Model Estimation and Interpretation

2.6.1 Dependent and Explanatory Variables

In this paper, we model the choice of the "primary" commute mode. Respondents selected their primary mode from a list of 13 alternatives in the survey. However, some of these alternatives were not chosen by many; others were conceptually rather similar. Accordingly, for the purposes of this study, we grouped the 13 alternatives into five broader categories: (a) driving alone, (b) shared ride (including carpooling, vanpooling and taking an employer shuttle), whether as driver or passenger, (c) local transit (bus, light rail, subway) – referred to as "transit" hereafter, (d) intercity/commuter rail, and (e) biking. Each respondent was asked to report perceptions for four of those five mode categories. Everyone was presented with categories (a), (b) and (c)[5]; in addition, online respondents whose one-way commute distance was less than 10 miles were asked to report mode perceptions for category (e), and everyone else (including all paper survey respondents) was presented with category (d). Although we allowed people to report walking as a primary commute mode and provide their perceptions for it, only 40 respondents in the full sample chose this mode. For this reason, we simplified the universal choice set by excluding those cases from the working dataset.

A respondent was assigned (for example) a transit alternative in his choice set if travel time and cost could be obtained, and if self-reported mode perceptions for bus, light

---

[5] For each of categories (b) and (c), respondents were instructed to choose a specific mode to rate, as follows: "Please answer with respect to ***ONE*** of the following means of transportation: the one you ***actually use most,*** if applicable, or else the one ***most realistic for your current commute circumstances.*** EVEN if you seldom or never use this means of travel, your responses will help us understand WHY you don't use it. ***Please check the box indicating which means of transportation you have in mind when answering these questions.***"

rail or subway were present. A total of 45 respondents were excluded from the working sample because they reported perceptions for only one mode; the remaining cases had 2-4 modes in their choice set.

As discussed in Section 2.4, the explanatory variables available for this study include the following:

- general attitudes, personality traits, and attitudes toward waiting;
- scores on the time use and preference factors shown in Table 2.3;
- perceptions of four modes, as shown in the "mode perceptions" block of Table 2.2, which were condensed into the three factors plus single item shown there;
- travel time and cost, which were externally obtained for each feasible mode using online tools including Google Maps and other sources (as described in Section 2.4), as well as headways for the transit and rail modes, using the same sources;
- the propensity to use a laptop while traveling, which was computed for each case using mode-specific binary logit models (as described in Section 2.5.2); and
- socio-demographic variables.

The mode perceptions, travel time (with the exception of in-vehicle travel time, which was allowed to have a different coefficient for biking), travel cost, and the propensity to use a laptop (travel multitasking) are generic variables (following Hensher and Johnson, 1981, we use "generic" to describe a variable that can take on different values for each alternative for a given person). The remaining variables are individual-specific, and they were given alternative-specific coefficients in the model estimation (with driving alone as the base alternative).

To reproduce population mode shares (Table 2.1) and remove the bias in coefficient estimators that would otherwise be associated with our essentially choice-based sampling strategy, each case was weighted (by the ratio of population to sample market shares for the alternative chosen by that person) in the calculation of the log likelihood function and resulting probabilities (Ben-Akiva and Lerman, 1985).

2.6.2 Model Results

Table 2.6 presents the summary statistics for the final mode choice model (estimated with NLOGIT 6) and its benchmarks. An alternative way of presenting the results of the estimation – as utility function equations –can be found in the appendix, Section A.3. The final model explains 58% of the information in the data, of which 90% is accounted for by the variables *other* than the alternative-specific constants in the model. This is considered quite good for a five-alternative revealed preference mode choice model. Consistency with the Independence of Irrelevant Alternatives (IIA) assumption was investigated by conducting Hausman-McFadden tests, and by evaluating a number of alternative model structures, including several nested logit and cross-nested logit specifications. All of these tests failed to reject the null hypothesis that IIA holds in this case. Since some of the modes would be considered "similar" to each other (notably commuter rail and transit; drive alone and shared ride; and shared ride and transit), this is a useful illustration of the point that IIA is a property that a given model specification may or may not have, and not a property inherent to a set of alternatives (Train, 2009). A well-specified model can capture among its *observed* variables the characteristics common to multiple modes, leaving its *unobserved* variables uncorrelated as is required for IIA to hold.

Coefficients for all the core generic variables have the expected signs and are strongly significant in the model. In the subsections below, we discuss key results for each group of variables in turn.

Table 2.6 – Multinomial logit commute mode choice model (weighted sample)

Variables	Biking	Commuter rail	Transit	Shared ride	Driving alone
			Socioeconomic characteristics[a]		
Driver's license	–[b]	–	–1.890** (0.8231)	–	base
Female	–	–	–	0.393*** (0.1510)	base
Race: white	–	–	0.532** (0.2167)	–	base
Limitation on walking	–	–	–	0.166*** (0.0559)	base
			Objective mode attributes		
In-vehicle travel time, min	–0.163*** (0.0592)	· · · · ·	–0.016***[c] (0.0056)	· · · · ·	· · · · ·
Out-of-vehicle travel time, min	· · · · ·	· · · · ·	–0.048*** (0.0089)	· · · · ·	· · · · ·
One-way commute cost, ln($)	· · · · ·	· · · · ·	–1.175*** (0.1375)	· · · · ·	· · · · ·
			General attitudes		
Pro-active modes	2.088*** (0.4029)	–	–	–	base
Pro-transit	–	0.954*** (0.2931)	0.825*** (0.1150)	0.201*** (0.0818)	base
			Multitasking preference		
Polychronicity	–	–	–	0.191*** (0.0693)	base
			Mode perceptions		
Mode convenience	· · · · ·	· · · · ·	0.455*** (0.0616)	· · · · ·	· · · · ·
Mode benefit /cost	· · · · ·	· · · · ·	0.368*** (0.0679)	· · · · ·	· · · · ·
Mode comfort	· · · · ·	· · · · ·	0.405*** (0.0563)	· · · · ·	· · · · ·
Mode multitaskability	· · · · ·	· · · · ·	0.098** (0.0431)	· · · · ·	· · · · ·
			Propensity for productive travel multitasking		
Propensity to use a laptop/ tablet/ netbook	· · · · ·	· · · · ·	1.240*** (0.3036)	· · · · ·	· · · · ·
			Constants		
Constant	–5.327*** (1.0289)	–2.959*** (0.3607)	0.785 (0.8272)	–2.752*** (0.2227)	base

Table 2.6 (continued)

Number of observations	2229	$\mathcal{L}(\hat{\boldsymbol{\beta}})$	–1127.247
$\mathcal{L}(\mathbf{0})$ (varying choice sets[d])	–2655.817	$-2(\mathcal{L}(\mathbf{0}) - \mathcal{L}(\hat{\boldsymbol{\beta}}))$	3057.140
$\mathcal{L}(\boldsymbol{c})$ (varying choice sets)	–1555.064	ρ^2	0.5756
$\mathcal{L}(\hat{\boldsymbol{\beta}})$ w/o constants	–1272.557	Adjusted ρ^2	0.5673

[***], [**] = significant at 1%, 5%.

[a] Effects of the variables are represented by an estimated coefficient and standard error (in parentheses).

[b] Dashes indicate coefficients that were constrained to be zero after they were found to have significance > 0.05.

[c] Centered coefficients preceded and followed by dots represent generic coefficients (i.e., constrained to be equal across the alternatives indicated by the dots).

[d] Note that when choice sets differ by individual, the unweighted equally-likely log-likelihood is not $-N \ln J$ as it is when all N cases have the same J alternatives, but rather $-\sum_n \ln J_n$, where $J_n \leq J$ is the number of alternatives in person n's choice set. This number will be larger (less negative) than $-N \ln J$, reflecting the information contained in the assumption that some alternatives have a zero probability of being selected rather than $\frac{1}{J}$ (a similar comment applies to the market-share log-likelihood, $\mathcal{L}(\boldsymbol{c})$). For comparison, the unweighted equally-likely log-likelihood corresponding to equal choice sets is $-2229 * \ln 5 = -3587.437$, so the difference is considerable. The *weighted* equally-likely log-likelihood for individual-specific choice sets is equal to $-\sum_n w_{j(n)} \ln J_n$, where *j(n)* is the alternative chosen by person *n* and $w_{j(n)}$ (defined as the population share of *j(n)* divided by the sample share of *j(n)*) is the weight for someone choosing alternative j. In the special case of equal choice sets, the unweighted and weighted equally-likely log-likelihoods are equal.

2.6.2.1 Socio-Demographic Variables and Objective Commute Mode Attributes

Several socio-demographic variables are significant in the model, all with expected signs. Except for *limitation on walking*, which is a standardized score (created from the 3-point ordinal responses to the item “Do you have any physical conditions or anxieties which prevent or limit you from walking?”), the rest of the socio-demographic traits are measured with dummy variables (= 1 if an attribute is present, and 0 otherwise). The possession of a driver’s license noticeably lowers respondents’ utility for transit as a commute mode. This means that to target drivers, public transit must outweigh that effect through superiority on other characteristics (e.g., including the ability to multitask). Gender differences appear only with regard to shared rides: females are more likely than males to carpool/vanpool or to take a shuttle. These results are consistent with others in the travel behavior literature. For example, Rosenbloom and Burns (1994), citing several national and international studies, point out that women are more likely both to carpool and to drive alone to work. These choices are often determined by family obligations, types of jobs available to females, household and work locations, prevailing income levels, single parenthood, etc. The more recent study by DeLoach and Tiemann (2012) corroborates this finding, specifically for fampooling (sharing a ride with a family member).

Respondents who identified themselves as white have higher utilities for public transit modes. This is a somewhat counterintuitive finding, because whites are the least represented race group among public transportation riders nationally (cf. AASHTO, 2015). The effect is probably associated with the local conditions in the study region of Northern California, where areas intensively covered by transit networks often overlap with areas having a higher prevalence of affluent white residents. Various types of physical and mental limitations are measured through the standardized variable *limitations on walking* (with a higher value corresponding to greater limitations). This variable is significant with a positive coefficient for shared ride. In other words, commuters who have stronger limitations on walking have a higher probability of sharing a ride with others than do those with weaker or nonexistent limitations.

The coefficients for in-vehicle and out-of-vehicle travel time (IVTT and OVTT) and for the natural logarithm of travel cost are negative, which is consistent with conventional wisdom. In the final model specification, we estimated two different IVTT coefficients, respectively for the biking alternative only and for all other modes. The estimated IVTT coefficient for biking is more than ten times larger in magnitude than the one for the other modes. In other words, according to the model results, the typical commuter in our sample would prefer 10.23 minutes commuting inside a vehicle over each minute on a bicycle, for example as an effect of the greater physical effort required by biking, or the effects of unpleasant traffic conditions, adverse weather, and topography. As the differences among the other alternative-specific coefficients for IVTT were not statistically significant, we constrained those coefficients to be equal across alternatives, for the sake of parsimony. OVTT, which is the sum of walking and waiting time for commuter rail and public transit alternatives, is perceived as three times more onerous than non-biking IVTT, a finding which is consistent with the dominant literature: public transportation users perceive access and waiting times as more inconvenient than in-vehicle travel time.

We tested multiple model specifications with different transformations of the one-way commuting cost variable. Allowing alternative-specific coefficients for this variable resulted in a counterintuitively insignificant coefficient for commuter rail, and having a non-log-transformed generic coefficient for the travel cost variable caused the generic coefficient of the propensity to use a laptop variable to become insignificant. However, besides resulting in strongly significant generic coefficients for the cost and propensity variables, a logarithmic transformation of the travel cost variable, a standard practice, produced a better fit to the data, and thus, it was selected as the final specification.

Monetizing the utility of saving travel time (via a VOTTS computation) has long been an important subject for transportation research and planning, and as indicated in Section 2.3, there has been considerable speculation about the impacts of travel-based multitasking on VOTTS. Accordingly, it is of interest to examine the VOTTS implications of the present model, to position it relative to more conventional mode choice models in the literature. We reserve a more in-depth investigation of the impact of multitasking on VOTTS for a separate paper (Malokin et al., 2017a; Malokin et al, 2018a).

The log-transformation of travel cost causes the value of travel time savings (VOTTS) to vary across individuals. Figure 2.2 summarizes the distribution of the VOTTS, respectively for IVTT and OVTT, among the commuters in the sample. The weighted mean VOTTS for non-bikers is 2.15 U.S. dollars per hour ($/hr) for IVTT and $6.45/hr for OVTT, respectively. The weighted median VOTTS of IVTT and OVTT for non-bikers is respectively $1.63/hr and $4.90/hr. At first glance, these estimates are substantially lower than the simple ballpark values suggested by U.S. DOT (35-60% and 80-120% of the hourly wage for IVTT and OVTT, respectively) (Trottenberg, 2011). According to the results of the model estimation, only 1.89% and 3.08% (respectively for IVTT and OVTT) of respondents have a weighted individual value of travel time savings that falls into the aforementioned ranges. However, Hensher and Wang (2016) showed that correcting for productive and leisure time while traveling for business purposes reduces VOTTS by 35%, 59%, and 42% for car, train, and bus respectively, supporting the expectation that conventional VOTTS numbers are inflated by the neglect of this factor. Furthermore, the literature identifies several sources that affect empirically-derived VOTTS in specific samples (in contrast to the simplified and policy-influenced guidelines set by government agencies): trip purpose, trip mode, distance traveled, travelers' income, etc. Applying the results of previous meta-analysis studies (Shires and de Jong, 2009 and Abrantes and Wardman, 2011) to our context, the VOTTS modeled as a function of these factors would yield a mean of $3.78/hr for IVTT and $4.81/hr for OVTT, after correcting for historical currency exchange rates and inflation. Thus, our results are largely in line with those in the academic literature. However, the model specification, namely including attitudinal and multitasking variables, also influences the computed VOTTS: these factors may lower the estimated VOTTS in our sample. For a more detailed discussion of VOTTS variability in this study, refer to Malokin et al. (2017a, 2018a).

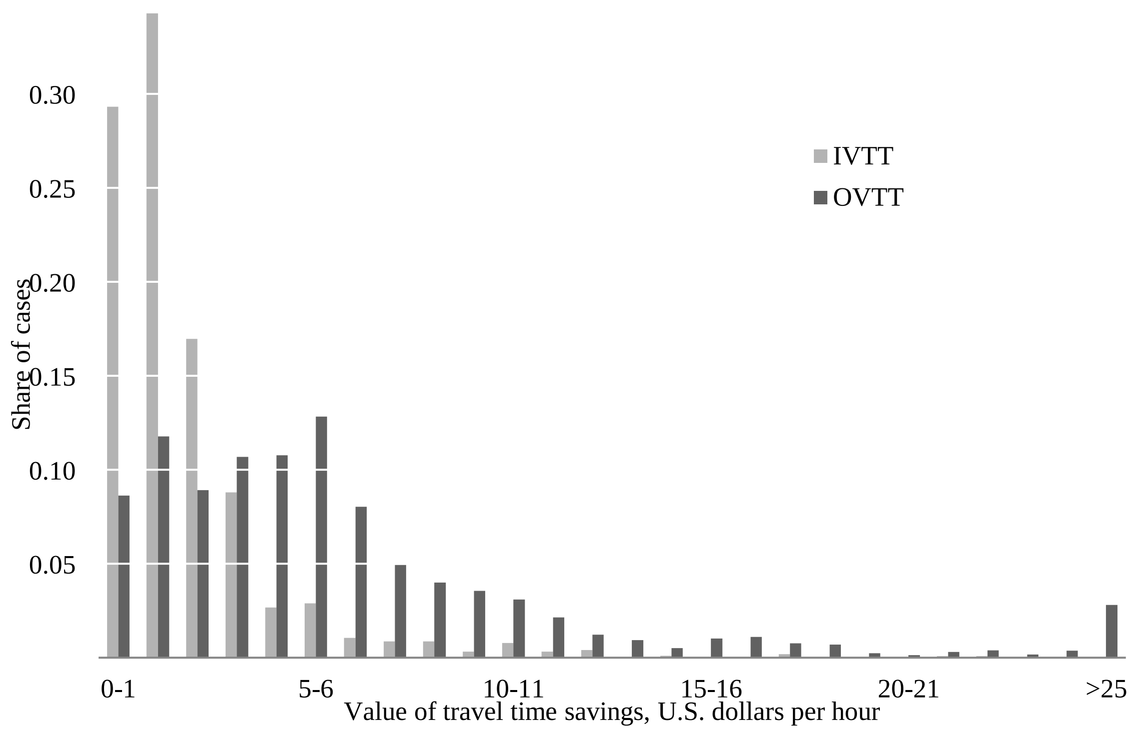

Figure 2.2 – Distribution of the weighted value of travel time savings (IVTT and OVTT) for non-bikers in the sample (N = 2037)

2.6.2.2 General Attitudes

Among the attitudinal variables that have a significant impact on mode choice, the *pro-transit* factor is positively associated with the choice of commuter rail, public transit, and shared ride modes. The *pro-transit* factor has a double connotation of conveying respondents' preferences both to take transit and to avoid driving as often as possible: with "driving alone" as the base, a positive factor score adds to the utility of the affected modes, to a greater extent for commuter rail, followed by transit and shared ride. This is a plausible result, given that commuter rail is likely to have a higher concentration of choice riders while local transit usually draws more captive riders (e.g., Giuliano, 2005; Taniguchi, 2012). The effect on shared ride users could be due to the half-way nature of this mode, with characteristics falling between those of the driving alone and public transportation modes.

Another "lifestyle" factor has significant effects on mode choice: the *pro-active* (non-motorized) *mode* attitude manifests people's desire to walk or bike instead of driving whenever possible. Not surprisingly, this factor is strongly significant for the bike alternative, with a large positive coefficient.

2.6.2.3 Mode Perceptions, Multitasking Preference, and Multitasking Propensity

All four mode perceptions have positive coefficients, meaning that the more favorably a given mode is perceived on various attributes, the greater the probability that it will be chosen. Since all four perceptions are standardized, by comparing the coefficient magnitudes we can note their order of importance to mode choice. *Convenience* has the greatest impact among them, followed respectively by *comfort*, *benefit/cost*, and the *ability to multitask*. Judging by the magnitude of the coefficients, it appears that a mode's convenience and comfort are respectively more than four and three times as important, in terms of effect on the perceived mode utility, as its multitasking conduciveness. However, given that the ability to multitask is measured as a standardized single item (rating of each mode on "ability to do things I need/want while traveling") while the remaining three perceptions are factor scores based on several items, it is reasonable to speculate that the multitasking variable has greater measurement error, and therefore that its coefficient will have a greater attenuation bias than the others (e.g., Cameron and Trivedi, 2005, Chapter 26). In any case, it is probably fair to say that although the ability to perform activities while traveling has a significant effect, it is not a dominant criterion in mode choice considerations. Both facets of this result are quite consistent with expectations. Further, given that the mean rating of the drive-alone mode on this item is higher than the mean ratings for other modes (both overall, and specifically for choosers of shared ride and bicycling as well as driving alone), it is also possible that a sizable share of respondents interpreted the item in a way differently than we intended[6], which could have additionally attenuated the variable's influence.

We also tested the impact of an individual's general propensity to multitask. This factor is significant only for the shared-ride mode (relative to driving alone). Its coefficient has the expected positive sign, although the magnitude indicates a rather small impact on the mode choice. This suggests that a general orientation toward multitasking is largely unrelated to one's mode choice consideration directly. At first glance, it might be

[6] We deliberately avoided the use of the term "multitasking" in this item, out of concern that not all respondents would be familiar with the term or think of "activities conducted while traveling" as multitasking. We also deliberately placed the item in question ("ability to do things I need/want while traveling") after the item "ability to run errands on the way to/from work" in the block of 14 perceived mode attributes, to minimize possible confounding of the two. However, it is possible that respondents were still prompted to think of the ability to make multiple stops along the way when they encountered the item in question. Alternatively, in the case of choosers of driving alone, it is possible that they subconsciously tailored the things they "need/want" to do while traveling to fit the capabilities of their preferred mode, or also, of course, that they truly did not want need or want to do things that were incompatible with driving.

expectable that highly polychronic people would be more likely than others to multitask while traveling; however, it is also plausible that monochronic individuals do not see the routine commute as a real distraction competing for their attention. Rather, a person's commute might be considered a passive background over which an active task can easily be laid (Circella et al., 2012). To the extent that both mechanisms are at work simultaneously, a polychronic orientation will have little to no influence on mode choice, on average. Still, individual propensities for conducting selected activities are a function of various polychronicity measures, as shown for laptop in Section 2.5.

Finally, the laptop propensity variable was also significant in the model. As described in Section 2.5, this variable was modeled using the subsamples of mode choosers, where the specifications (including the alternative-specific constants ensuring that mode-specific shares of laptop adoption were replicated) of those mode-specific models already accounted for the differential conduciveness to using a laptop that pertained to each mode. Thus, driving alone and transit tended to have low conduciveness, while commuter rail had high conduciveness. This result is natural: for drivers, the mental and physical resources required for the driving task prevent the individual from efficiently performing complex activities while driving. For transit users, we can speculate that lower comfort on board (including crowding and vibrations) and the potential subdivision of a trip into several shorter legs (thereby increasing the "overhead" time of unpacking and packing one's productivity tools; see, e.g., Watts, 2008) creates an unsupportive environment for productive tasks. In contrast, commuter rail (at least in the area of study) usually offers a seat, tables, electric outlets, internet connectivity (Wi-Fi), and longer (on average) in-vehicle travel time, all of which contribute to increasing the propensity to be productive while traveling.

The estimated positive coefficient for this variable indicates that a propensity to be productive "on the go" while commuting on a certain mode *increases* the utility of that mode. When model specifications with alternative-specific coefficients for the laptop propensity variable were tested, only commuter rail and shared ride had statistically significant estimated coefficients (positive, as were the insignificant coefficients for drive-alone and transit, with the one for biking set to zero since the variable does not vary within that alternative). However, using alternative-specific coefficients presents a potential conceptual challenge. Although in principle we saw nothing wrong with allowing a unit of propensity to have a different *impact* on the utilities for different modes, in addition to the propensity *itself* already taking on a differential value for different modes (just as travel time might have an alternative-specific *coefficient* as well as an alternative-specific *value*), it seemed unnecessarily cumbersome, conceptually, to do so. For one thing, our propensity variable differs from travel time in that time *does* have an objective measurement, and allowing time to have an alternative-specific coefficient reflects an assumption that the *perception* of the (dis)utility of time could differ by mode. By contrast, the propensity variable, itself a function of the utility of using a laptop, is by nature subjective at the outset, and, as mentioned, by nature already accounts for the mode-specific differences in that utility. For another thing, allowing the coefficient as well as the value of the propensity variable to differ by alternative would complicate the scenario-testing discussed in the next section: to reflect an autonomous vehicles scenario, should the *value* of the propensity variable be manipulated, its *coefficient*, or both? Accordingly, we retained laptop propensity as a generic variable in the model.

2.7 Role of Multitaskability: Transit-Advantage and Autonomous Vehicle Scenarios

Using the results from the model estimation, we can further investigate the contribution to commute mode choice of the propensity to use a laptop during the trip. We do so by presenting several what-if scenarios in which we estimate the changes in the weighted mode shares associated with different hypothetical values of the propensity to use a laptop and the multitasking ability perception. Table 2.7 summarizes the results of this analysis.

All scenarios are generated using various values of the *propensity to use a laptop* variable, with some scenarios also having altered *multitaskability mode perceptions* to reflect hypothetical "objective" changes in a mode's multitaskability. For simplicity, we do not assume any changes in other variables; i.e., essence we adopt the perspective of, "what would happen if laptops were unavailable, or AVs were available, *today*?" Except for the first scenario, they are grouped in pairs and denoted by Roman numerals. The first ("I") scenario in each pair is the result of manipulations only in the *propensity to use a laptop* variable, whereas the second scenario ("II") involves simultaneous changes in both the *propensity to use a laptop* and the *multitaskability mode perception* variables.

The first scenario, ***laptop unavailability***, simulates the potential mode shifts that would happen if the productive multitasking propensity did not have any effect on the utility functions of the alternatives. Conceptually, this scenario allows evaluating the mode shares that would be observed if laptops were not available (i.e., passengers were not able to carry out productive activities while traveling), or, in other terms, what proportion of mode shares are attributable to the ability to engage in this multitasking behavior.

The second and third scenarios, which both simulate ***rail dominance*** conditions, identify some possible *upper* bounds for commuter rail, i.e., a "maximum" attractiveness that commuter rail could have among commuters. In these scenarios, for each individual the rail-specific *propensity to use a laptop* is set to one (implying that all rail passengers would use laptops) and (for the II scenario) the *multitaskability commuter rail perception* is set to the highest value that variable takes on across the entire sample. Such universal appeal of productive multitasking will presumably never be achieved; however it is useful to assess such a ceiling: this represents the maximum share that commuter rail could gain, all else equal, based on its laptop usability advantage alone.

The next pair of scenarios, ***transit improvement***, looks at changes in mode shares associated with the enhancement of the same set of variables for the transit (local/express bus, light/metro rail) alternative. However, instead of setting the *propensity to use a laptop* on transit and the *multitaskability perception of transit* to the *highest possible* values (1 and the sample maximum, respectively), as in the previous two scenarios, in these scenarios we use the individual's currently observed variable values for *commuter rail* as a target (if the values for these variables are greater for commuter rail than for transit for that person, which is the case for 34% and 94% in the weighted sample for the propensity and perception variables, respectively). By doing so, we set a more realistic (albeit still ambitious) goal for the transit mode's conduciveness towards multitasking, considering that commuter rail is, on average, objectively and subjectively superior to transit in this respect (e.g., due to seat availability, and the presence of tables, electric outlets, and internet connectivity). Effectively, these scenarios evaluate how popular transit services could be if they were perceived as offering the same opportunities to use a laptop as commuter rail.

Finally, we assess the changes in commute mode choice associated with the multitasking potential of fully autonomous vehicles (AVs). Among many other revolutionary impacts, AVs (at the final level of automation, as commonly classified by NHTSA, 2013) can expand the set of activities that can be conducted while operating the vehicle well beyond what is currently feasible while driving cars. Accordingly, in the ***autonomous vehicles*** scenarios, for each person we assign the drive-alone and shared-ride modes the same levels of travel-based multitasking "conduciveness" as she currently perceives for commuter rail (if the latter is larger[7]). For example, if a person has a propensity of 0.7 to use a laptop on commuter rail, we assign this value to the driving-alone and shared-ride propensities (if they are smaller), simulating an AV-level of multitaskability of the latter modes. Of course, a personal (autonomous) vehicle will not be exactly equivalent to commuter rail, even in its multitaskability. On the one hand, rail may offer a ride less susceptible to motion sickness, and may offer distinctive amenities such as toilets and diverse food options. On the other hand, personal vehicles offer privacy (potentially including portable toilets), greater insulation from "stranger danger", mobility (no transfers, real-time route adjustment, etc.), and ownership (including the potential for storage of productivity tools, and storage/preparation of food). Considering the pros and

[7] The rail-specific multitasking propensity (target) is greater than the *shared ride*-specific propensity for 48% in the weighted sample, while the sample maximum multitaskability perception (target) is higher than the *shared ride*-specific one for 95% in the weighted sample. Respectively, the targets are larger than the *driving alone*-specific variables for 74% and 65% in the weighted sample.

cons of each mode, commuter rail arguably provides a reasonable approximation to a travel multitasking environment for personal autonomous vehicles.

In Table 2.7, the AV scenarios have the implicit qualifier “full availability”. This refers to the assumption that everyone in the sample has AV alternatives in their choice sets. This approach constitutes the most extreme situation, or an upper limit. A more realistic approach would account for a gradual diffusion of AV technologies. In particular, as a simplification we can consider those individuals having the highest propensity to be productive during their commute (i.e., to *use a laptop/ tablet/ netbook*) to be among the earliest group to consider AVs, followed at a later time by the less avid technology users. We reflect this assumption by repeating the ***autonomous vehicles diffusion I*** simulation nine more times, varying the percent of cases having AVs available from 10% to 90%, where cases are first ranked on the basis of their highest propensity to use a laptop on commuter rail (and, therefore, on future AVs), and then, at each stage, the *x* percent of cases assumed to have AVs available to them are the *x* percent highest-laptop-propensity cases. Thus, in Figure 2.3, the *x*-axis represents deciles of adopters such that the first quantile (the earliest adopters) in the graph corresponds to the top decile of the individual multitasking propensity in the sample (the most avid laptop users). The *y*-axis shows the relative change in mode shares given the availability stage. According to Figure 2.3, the bulk of the mode share shift has happened by the time the 40% of commuters with the greatest multitasking propensities consider AVs.

Table 2.7 – Weighted mode shares under various assumptions on multitasking propensity/ mode multitaskability, %

Scenario name[a]	Assumptions	Biking	Commuter rail	Transit	Shared ride	Driving alone
Current population shares		1.534	0.716	8.174	12.460	77.117
Laptop unavailability	For each alternative, *propensity to use a laptop* set to zero.	1.566 0.033	0.609 –0.110	7.940 –0.234	11.282 –1.178	78.605 1.488
Rail dominance I	For commuter rail alternative, *propensity to use a laptop* set to one.	1.534 0.000	1.444 0.727	8.113 –0.061	12.294 –0.165	76.615 –0.501
Rail dominance II	For commuter rail alternative, *propensity to use a laptop* set to one and *multitaskability perception* set to maximum for that alternative.	1.534 0.000	1.589 0.872	8.102 –0.072	12.264 –0.195	76.511 –0.605
Transit improvement I	For transit alternative, *propensity to use a laptop* set equal to the *propensity* for commuter rail (if greater).	1.526 –0.008	0.703 –0.013	8.611 0.437	12.347 –0.113	76.813 –0.303
Transit improvement II	For transit alternative, *propensity to use a laptop* set equal to the *propensity* for commuter rail (if greater) and *multitaskability perception* set to the sample maximum for the commuter rail alternative.	1.510 –0.024	0.695 –0.021	9.329 1.155	12.205 –0.255	76.261 –0.855

Table 2.7 (continued)

Scenario name[a]	Assumptions	Biking	Commuter rail	Transit	Shared ride	Driving alone
Autonomous vehicles diffusion (full adoption) I	For shared ride and driving alone alternatives, *propensity to use a laptop* set equal to the *propensity* for commuter rail (if greater).	1.475 –0.059	0.594 –0.122	7.707 –0.467	11.829 –0.631	78.395 1.279
Autonomous vehicles diffusion (full adoption) II	For shared ride and driving alone alternatives, *propensity to use a laptop* set equal to the *propensity* for commuter rail (if greater) and *multitaskability perception* set to the sample maximum for the commuter rail alternative.	1.393 –0.141	0.531 –0.185	7.157 –1.016	12.323 –0.137	78.596 1.479

[a] The first row in each band displays mode shares expressed as percentages, and the second row presents the change in percentage points from the current shares under each scenario. Note that at the time of this writing, the simulation function of Limdep 10.0 does not account for weights in its calculation of the predicted aggregate alternative shares. The weights have been incorporated into the results shown here.

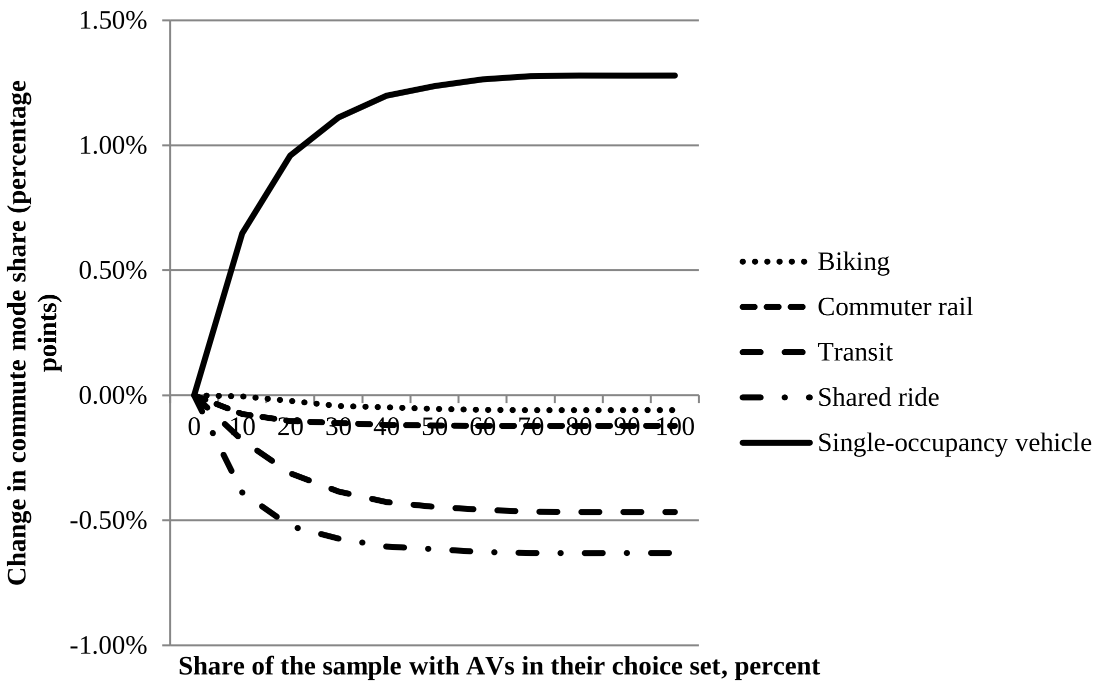

Figure 2.3 – Mode share changes as a function of AV availability in people's choice (consideration) sets, assuming availability is associated with the propensity to use laptop

In the ***laptop unavailability*** scenario, all modes lose market share to driving alone (which shows an increase of 1.5 percentage points) and, marginally, biking. The most disadvantaged mode is shared ride, which drops by 1.2 percentage points (p.p.), followed by transit (0.2 p.p.) and rail (0.1 p.p.). Reversing the base of this scenario, we can evaluate these market share changes as the effect attributable to the use of a laptop during the commute. In other words, the mode share for driving alone would be greater by 1.5 p.p. if people had no propensity (ability) to use a laptop while traveling. A similar logic applies to the increase in ridership of the collective modes, thus giving a measure of the positive impact that travel-based productive multitasking has had on the popularity of some travel modes.

The ***rail dominance*** scenarios exhibit increases in the mode share for commuter rail, which would be gained at the expense of all other alternatives (except biking, which is virtually unaffected). Still, while the increased behavioral propensity to multitask on commuter rail is responsible for a substantial increase in mode share (0.7 p.p.), the incremental impact of the improved perceptions was more modest (0.15 p.p.). This could be a sign of the favorable recognition among travelers of this alternative, which is already perceived as very conducive to productive multitasking. Under these conditions, an increased propensity to use a laptop (specifically a scenario in which everybody on board uses a laptop) still has some potential to attract additional riders. It is also of interest to note the substitution patterns across modes in the ***rail dominance*** scenarios. Namely, the main "donor" of the increased mode share for commuter rail is driving alone, which loses 0.5 and 0.6 p.p. of mode share in the two scenarios, respectively. By contrast, mode shares for transit have very modest decreases of 0.06 and 0.07 p.p. This seems to suggest different natures for the ridership bases of these public transportation modes.

The ***transit improvement*** scenarios have virtually no effect on biking and commuter rail, while an increase is found in the transit mode shares (+0.44 and +1.16 p.p.) at the expense of shared ride (–0.11 and –0.26 p.p.) and, more profoundly, driving alone (–0.30 and –0.86 p.p., respectively). Differently from the ***rail dominance*** scenarios, an increase in the perception of travel multitasking ability for transit has a substantial impact on its mode share. This could signify the importance for local transit operators of providing an environment that is more favorable towards multitasking if the main priority is to attract new riders.

Finally, according to the ***autonomous vehicles diffusion*** scenarios, an increase is expected in the mode share for single occupancy autonomous vehicles (SOAVs) (1.28 and 1.48 p.p., respectively, in the two scenarios). Interestingly enough, the shifts in the mode share for shared-ride AVs (SRAVs) do not show the usual compounding effect that was observed in the other scenario pairs, i.e., the magnitude of the change does not further increase in the second scenario (–0.14 p.p.) compared to the first (–0.63 p.p.). Therefore, improving the multitaskability perception of SRAV partially mitigates its inferiority to SOAV, but all else equal, SRAVs still have a less multitasking-"friendly" environment compared to SOAVs. The main source of the increased demand for automobile-based modes under the ***autonomous vehicles diffusion*** scenarios is transit, followed by commuter rail and biking, which reinforces a common notion that automation technology is expected to damage public transportation.[8] It is worth stressing again that the shifts in mode shares predicted in this study are only associated with some aspects of a given mode, namely its multitaskability. They are oblivious to the other disruptive effects of automation technology on transportation demand.

A key concern with autonomous vehicles is the potentially great increase in the number of cars on the road, exacerbating traffic congestion and pollutant emissions. The ***autonomous vehicles diffusion (full adoption) II*** scenario shows that the mode share for driving alone would increase by 1.5 p.p., and the shared ride share would decrease by 0.14 p.p., holding all else constant (e.g., the costs of owning and maintaining a personal vehicle, travel times, etc.). These results translate into almost 59,000 additional vehicles per day on

[8] In this study, we do not model any automation of public collective modes, as the impact of such automation on the ability to multitask while commuting may be negligible.

the roads in the study area's commute shed (assuming an average vehicle occupancy of 2.42 people for car/vanpools in the study area), where currently 4.12 million daily car commutes are made (according to the 5-year American Community Survey data, 2006–2010). However, additional changes in vehicle ownership (e.g., increased popularity of carsharing options that involve calling a driverless car only when needed) and intelligent dynamic rideshare matching could be possible solutions, among others, that might at least partially curb the infrastructure load.

2.8 Conclusions and Future Research

In this study we investigated the impact of multitasking attitudes and propensities on mode choice and valuation of travel time – to our knowledge, the first revealed preference model developed for this application. Based on a survey of 2229 Northern California commuters, we built an MNL model to predict commute mode choice as a function of objective mode characteristics, socio-economic aspects, individual attitudes and traits, time use, and activities conducted while commuting. Although engagement in activities was reported only for the chosen mode and therefore could not be directly used in the MNL model, we circumvented this endogeneity bias by estimating binary logit models of the propensity to conduct each activity while traveling on each mode, and tested the inclusion of the predicted propensities as explanatory variables in the mode choice model. We selected the *propensity to use a laptop/ tablet/ netbook* as a measure of the inclination to engage in a major type of productive travel-based multitasking, to use in our final mode choice model.

The MNL estimation results show that multitasking is significant to mode choice in three ways. A *generic mode perception* coefficient is positive, indicating that greater perceived multitaskability of a travel mode adds to its utility (and, therefore, to the likelihood that the mode is chosen). The alternative-specific (shared ride) coefficient for the *general preference towards multitasking* (polychronicity) is also positive and strongly significant. And the generic coefficient for the *propensity to use a laptop/ tablet/ notebook* (i.e., propensity for this form of productivity during travel) is positive and strongly significant.

We generated a set of scenarios, through manipulating the mode-specific multitaskability perception and propensity to use a laptop while traveling on several modes, to: (a) estimate the effect of the positive utility added by the current level of engagement in productive activities (using laptop); (b) assess the potential of travel-based multitasking to increase the appeal of commuter rail and transit at the expense of other modes; and (c) evaluate the potential impacts attributable to the adoption of autonomous vehicles, when they will offer the same level of multitasking "conduciveness" that commuter rail offers today. In view of the current existence of other ways to be productive while traveling, as well as the continued evolution of technology to permit more and better ways to be productive in the future, the impacts of *propensity to use a laptop/ tablet/ netbook* identified by this study could be viewed as a conservative estimate of the total commute mode choice impacts of productivity-oriented travel multitasking.

Based on the findings of this study, the outlook for public transportation operators is mixed: in the short and medium term, the transit improvement scenarios of Table 2.7 suggest that public transit modes (here defined as local/express bus and light/metro rail) have the potential to increase their ridership by appealing to the productivity attitudes and behavior of commuters. For example, a recent study (Dong et al., 2015) estimated that introducing free Wi-Fi service increased ridership on Amtrak's Capitol Corridor (California) intercity train service by 2.7 percentage points. However, in the long term public transportation might be threatened by the diffusion of autonomous vehicle technology, which may ultimately attract commuters by providing superior mobility and accessibility (potentially with reduced costs of ownership and operation achieved through carsharing and/or intelligent rideshare matching), and with increased ability to use travel time productively. This might reinforce the existing tendency towards low-density development, further propagating urban sprawl (e.g., Mokhtarian, 2018).

The laptop propensity and mode choice models were estimated on a diverse commuter cohort which is, we believe, illustrative of urbanized populations in the United States. Clearly the cultural, transportation, and urban form contexts will be different elsewhere in the world, and thus it is unlikely that the specific empirical results of this study will be replicated outside the U.S. The methodology, on the other hand, should be widely transferable – not only to other geographic contexts, but to other dimensions of the commute mode choice context and to different choice contexts altogether. The key features of the methodology include (1) the incorporation of pertinent attitudinal variables as well as more traditional attributes of the alternatives[9] and socio-economic traits; (2) the existence of a consequence of interest (productive multitasking, in our case) that is only observed conditional on the main choice of interest (primary commute mode, in our case)

[9] In the present application to mode choice, an additional feature of the methodology is the collection of mode attributes via Google/ Bing Maps API.

but (the anticipation of) which is assumed to influence that main choice; (3) the estimation of "consequence adoption" models on choosers of each alternative, followed by the application of those models to the entire sample to generate predicted propensities to experience that consequence if each of the given alternatives were to be chosen; (4) the incorporation of those predicted propensities as explanatory variables in a model of the main choice; and (5) plausible manipulation of those predicted propensities and/or other explanatory variables, applied to the previously-calibrated main choice model to estimate the impacts of various hypothetical scenarios. As an example of a different choice context, this methodology could potentially be applied when the main choice of interest is the adoption of shopping mode or channel (store versus online) and the consequence of interest is whether the item needs to be returned.

Related studies underway or recently completed using this same dataset include investigations of the systematic heterogeneity in the disutility of travel time (Etezady et al., 2019) and of the reported benefits and disadvantages associated with the activities conducted while commuting (Shaw et al., 2018). The early results emerging in the present paper and in these other studies suggest that this will continue to be a worthy subject of investigation for some time to come, with much remaining to be learned as technology continues to alter the landscape of possibilities.

2.9 Acknowledgements

A previous version of this paper (Malokin et al., 2015) was presented at the 94th Annual Meeting of the Transportation Research Board (TRB). The thoughtful and constructive comments of five anonymous reviewers have greatly improved the present version. This project was initially funded by the Sustainable Transportation Center of UC Davis, which receives funding from the U.S. Department of Transportation and Caltrans, the California Department of Transportation, through the University Transportation Centers program. Additional financial support was provided by a new faculty grant from the Georgia Institute of Technology and by a partnership with the Capitol Corridor Joint Powers Authority (CCJPA). The authors are immensely grateful to Amanda Neufeld and Zhi Dong, who were deeply involved in the survey design and organized the data collection and cleaning effort. We wish to thank Javier R. Salgado, who helped with descriptive analysis and additional data cleaning. Special gratitude goes to Taha Rashidi, who suggested the multidimensional nested logit modeling of joint choices of mode and laptop usage. Cheng Zhuo, Aurina Lam, Eileen Coleto, Adam Stocker, Valerie Onuoha, Andre Tu, and Kelly Caines provided vital assistance during the data collection and entry. Additionally, we express our gratitude to Chandra Bhat, David van Herick, Hani Mahmassani, Aruna Sivakumar, Laurie Garrow, and William Greene for useful discussions on choice-based sampling and other technical matters.

CHAPTER 3. VALUE OF TRAVEL TIME AMONG MULTITASKING MILLENNIALS

Malokin, Aliaksandr, Giovanni Circella and Patricia L. Mokhtarian (2018a) Do multitasking millennials value travel time differently? A revealed preference study of Northern California commuters. Available from the authors.

3.1 Abstract

Millennials, the demographic cohort born in the last two decades of the 20th century, are reported to adopt information and communication technologies (ICTs) in their everyday lives, including travel, to a greater extent than older generations. As ICT-driven travel-based multitasking influences travelers' experience and satisfaction in various ways, millennials are expected to be affected at a greater scale. Still, to our knowledge, no previous studies have specifically focused on the impact of travel multitasking on travel behavior and the value of travel time (VOTT) of young adults. To address this gap, we use an original dataset collected among Northern California commuters (N=2216) to analyze the magnitude and significance of individual and household-level factors affecting commute mode choice. We estimate a revealed preference mode choice model and investigate the differences between millennials and older adults in the sample. Additionally, we conduct a sensitivity analysis to explore how incorporation of explanatory factors, such as attitudes and propensity to multitask while traveling, in mode choice models affects coefficient estimates, VOTT, and willingness to pay to use a laptop on the commute. Compared to non-millennials, the mode choice of millennials is found to be less

affected by socio-economic characteristics and more strongly influenced by the activities performed while traveling. Young adults are found to have lower VOTT than older adults for both in-vehicle (15.0% less) and out-of-vehicle travel time (15.7% less), and higher willingness to pay (in time or money) to use a laptop (31-455% more), even after controlling for demographic traits, personal attitudes, and the propensity to multitask. This study contributes to better understanding the commuting behavior of millennials, and the factors affecting it, a topic of interest to transportation researchers, planners, and practitioners.

3.2 Introduction

The impact of activities conducted while traveling, i.e., travel-based multitasking, has become an emerging topic in travel behavior research in recent years. On one hand, the increased availability of portable and affordable information and communication technology (ICT) devices – including smartphones, tablets, laptops and, most recently, wearables – has prompted the research community to evaluate the significance of multitasking to transportation. ICT plays a constructive role in many areas of transportation: for example, affecting trip-making (e.g., overall trip generation, and the specific time of departure), trip experience, mode choice, and some travel characteristics. On the other hand, sustainability and other goals, coupled with the limited financial resources (and political will) available to meet them, have raised high expectations for modest-scale interventions that can help meet such strategic goals, at least partially. As one such intervention, travel-based multitasking promises to make travel time less onerous, and more productive and enjoyable. Possible effects could include an increased appreciation of the travel experience (a factor that can increase the number of users that are willing to use and/or pay for some modes and services), and measurable changes in mode choice, e.g., switching to public transit where conducting certain activities during a trip is more feasible. Longer-term effects might include changes in residential location and land use, if some individuals are willing to live farther from habitual destinations and do not mind the longer time spent travelling if this time is perceived as less wasted. Alternatively, of course, effects could be negative both personally (such as a perceived "contamination" of previously "private" travel time with expectations for accessibility and productivity) and societally (such as diminishing the disutility of travel time in automobiles as well as in transit vehicles).

Many opportunities to multitask, independent from the use of ICT devices, have been available to travelers for a long time. More recently, ICT-enabled multitasking has become a common feature associated with the increased availability of modern digital devices. As with many innovations, younger generations are among the early adopters. Accordingly, to the extent that ICT-related travel-based multitasking affects travel behavior, larger impacts can be expected on the travel choices of current young adults, or millennials, commonly dubbed "digital natives".

The millennial generation encompasses those who were born between, approximately, 1980 and the end of the 20th century. Millennials are currently the most populous cohort in the United States, and they have long been attracting considerable attention in consumer and travel behavior research. From the transportation standpoint, this generation is observed to have lower car ownership, lower per-capita vehicle-miles traveled (VMT), increased interest in urban residential locations, and higher adoption of digital technologies and shared mobility services (Blumenberg et al, 2012). As time progresses, millennials will play a defining role in shaping the travel patterns of the whole society. However, there is some evidence that, as millennials age, their travel behavior is converging with that of older generations (Garikapati et al., 2016). All of this makes the millennial cohort of particular interest to current travel behavior research.

Thus, this study combines three timely topics: we (1) analyze the impacts of activities conducted while traveling, and in particular the role of activities that rely on the use of ICT; while (2) investigating the transportation choices of millennials, and how they differ from those of previous generations; and (3) studying them through the lens of the value of travel time[10] (VOTT) and willingness to pay (WTP) for travel multitasking. Our analysis of the travel behavior of millennials and non-millennials is based on a rich dataset that we collected in 2011-2012 from Northern California commuters (the total sample includes more than 2,000 respondents). We estimate revealed preference mode choice multinomial logit (MNL) and artificial nested logit (ANL) models, segmented by respondents' age cohort, and discuss the impacts of commute and individual characteristics on the utility associated with each available mode.

Several research questions are addressed in this study. How and to what extent does the travel behavior of millennials differ from that of older generations? What variables influence the mode choices of the members belonging to each cohort? Do millennials have different values of travel time and willingness to pay for the ability to conduct activities while commuting (in particular, to use a laptop / tablet)?

[10] Following the discussion in Daly and Hess (2019), we opt for the term of "value of travel time" instead of "value of travel time savings".

The remainder of the paper is organized as follows. In Section 2 we review the existing literature and summarize previous research that focuses respectively on the analysis of the impact of activities conducted while traveling, the value of travel time estimation, and the travel behavior of millennials. Section 3 describes the study sample, including differences between millennials and non-millennials. The next section presents the mode choice models, and discusses the main significant explanatory variables. In Section 5, for both segments, we calculate the VOTT and the WTP for using a laptop / tablet, treating the latter as illustrative of ICT-based productive activities conducted while traveling. Section 6 presents a sensitivity analysis for WTP and VOTT. In the final Section 7, we conclude by discussing the findings and their significance for present and future travel demand.

3.3 Literature Review

This paper aims to bring together three topics: the impact of activities while traveling on mode choice, the estimation of WTP and VOTT, and the analysis of the travel behavior of millennials. Each of these areas has its own stream of well-developed literature that rarely intersects with the others. To the best of our knowledge, no study has investigated all three of these subjects together. Therefore, in the remainder of this section, we briefly highlight some previous research on each of these topics, respectively.

3.3.1 Impact of Activities while Traveling on Mode Choice

The research community's interest in the influence of travel-based multitasking on travel behavior has been on the rise throughout the past several years. The conceptual grounds of the impact of multitasking were laid down by DeSerpa, (1973), Lyons and Urry

(2005), Watts and Urry (2008) and Gripsrud and Hjorthol (2012). Several empirical studies have followed, focusing in particular on (1) the patterns of activities performed while traveling (e.g. Ohmori and Harata, 2008); (2) the number of activities performed while traveling, controlling for socio-economic and mode attributes (e.g. Zhang and Timmermans, 2010); (3) the impact of travel-based multitasking on the subjective evaluation of trip experience (Ettema et al., 2012; Susilo et al., 2012; Rhee et al., 2013; Rasouli and Timmermans, 2014; and Mokhtarian et al., 2015); and (4) the influence of a multitasking-friendly travel experience on mode choice (Zheng et al., 2016; Malokin et al., 2019). Keseru and Macharis (2017) have compiled a comprehensive review of travel multitasking studies to date.

Recently, Frei et al. (2015) used data from 336 Chicago-area riders to model the engagement in multitasking activities while riding public transit. Among several interesting findings, travel-based multitasking was found to be associated with a better travel experience by allowing saving time and increasing pleasure (e.g., reading a book). Further, an activity (e.g., listening to music) can be used to keep a passenger’s mind off the trip.

In our previous work, similarly to Frei et al. (2015), Berliner et al. (2015) analyzed the factors behind the engagement in travel multitasking, separately by the distinctive modes that are used, and distinguishing productive from hedonic, and ICT-based from non-ICT-based activities. Building on the findings in Berliner et al. (2015), the authors of the current paper built a revealed preference MNL mode choice model where individual-specific travel multitasking propensities were found to have significant, albeit modest, effects on mode choice (Malokin et al., 2019).

To the best of our knowledge, there has been no research that has specifically focused on the travel-based multitasking behavior of millennials. However, several studies that investigated various dimensions associated with travel multitasking used age as a predictive factor. For example, Frei et al. (2015) showed that young adults used ICT devices more actively than older transit users. Susilo et al. (2012) found that young adults of ages 16-25 years were more likely to use their time beneficially while traveling on a train, and consider their commute as wasted time. Mokhtarian et al. (2015) showed that the French Millennial generation was more inclined to evaluate their trips as mentally and physically tiring, and unpleasant, than older generations.

The influence of the activities conducted while traveling on VOTT has also been little studied (although often speculated), to date. We know of only two papers (Ettema and Verschuren, 2007 and Varghese and Jana, 2018) that have explicitly tackled this relationship empirically. In a stated preference study, Ettema and Verschuren (2007) demonstrated that age and polychronicity (preference for multitasking) influenced VOTT: younger travelers (unexpectedly) had higher VOTT (by 78%) than older ones; and the VOTT of monochronic commuters was 32% higher than that of polychronic ones, suggesting that multitasking preference decreased VOTT. The type of activities conducted while traveling also played a role in determining the VOTT: listening to music tended to lower it by 69%, while reading increased VOTT by 36%. Ettema and Verschuren (2007) illustrated the importance of travel-based multitasking in VOTT calculations by a sensitivity analysis. Accounting for monochronicity could increase VOTT by as much as 351%; and factoring in activities while traveling additionally changed VOTT between –59% and +46% compared to the base model with no multitasking effects. However, the applicability of the study results is constrained by a number of factors: (1) the limited number and the technological scope of the presented activities, (2) the stated preference study design, and (3) the inability to evaluate possible mode shifts.

In the other study, Varghese and Jana (2018) employed revealed preference trip diary data to estimate conventional (i.e., not containing attitudes) MNL and mixed MNL models separately for travel multitaskers and their opposites. Comparisons of the resulting VOTT estimates were used to quantify the effects of travel multitasking. They found that on average, multitasking reduces VOTT by 26%, while some of the activities, such as eating and listening to music, are associated with higher VOTT.

3.3.2 Variations in Value of Travel Time Savings Estimates

Willingness to pay (WTP) measures a marginal rate of substitution between two attributes. The VOTT is a special case of willingness to pay – measuring the substitution rate between travel time and travel cost – which is widely used in numerous economic applications that involve evaluating, predicting, and improving the effectiveness of transportation systems in channeling goods and people. Unsurprisingly, the extensive set of attributes of the system itself, the actors involved, and the relationships among them define quite a range of variations in the estimated VOTT. A few extensive meta-analyses and literature reviews have been published in the literature, identifying the main sources of variation in VOTT. Table 3.1 summarizes the most common causes of variability in VOTT.

In particular, two meta-analyses (Gunn, 2001; Hensher and Wang, 2016) found that conducting productive activities while traveling reduced VOTT. Both studies considered only business trip purposes and specifically excluded commuting and personal trips from the estimation of the impact on VOTT of productive activities while traveling. Gunn (2001) attributed a 23% decrease in VOTT between 1988 and 1997 among Dutch train business travelers to the improved in-car experience and diffusion of "laptop-computers", whereas Hensher and Wang (2016) estimated that productive and leisure activities while traveling lowered VOTT for business travelers by 35%, 59%, and 42% for car, train, and bus respectively.

In an attempt to evaluate VOTT as the difference between the opportunity cost of time (i.e., the cost of forgoing activities that compete with traveling for time) and any utility accruing from the time spent on activities while traveling, Kouwenhoven and Jong (2018) used a 2009-2011 sample of 822 Dutch travelers to analyze their stated preference responses on the perceived trip time usefulness with respect to changes in travel time and availability of ICT devices during travel. Expectedly, they found that travelers who find trip shortening useful (i.e., who place a greater value on activities outside the trip) had a VOTT 15% higher than those who did not. Similarly, those who reported being able to spend their travel time usefully had a 14-26% lower VOTT. Unexpectedly, however, they also found that travelers who had a mobile phone, computing device, or music player available to them exhibited a 10-20% increase in VOTT over those who did not have these ICT devices. The authors attempted to explain such a counterintuitive result by suggesting that travelers with ICT might be of higher income and busier.

Historically, the range of estimated VOTT has been wide. For example, in a comparative study Zamparini and Reggiani (2007) reported the range of VOTT, measured as a percentage of the wage rate, to be as low as 13% (Talvittie, 1972) and as high as 145% (Guttman, 1975) for commuting by car in the U.S[11].

Table 3.1 – Causes of variability in VOTT

Variable		Influence on VOTT … Positive	… Negative
Trip purpose	*Business*	Arbantes and Wardman, 2011 Zamparini and Reggiani, 2007	
	Commute	Arbantes and Wardman, 2011	
	Leisure		Arbantes and Wardman, 2011
Trip mode	*Bus*		Arbantes and Wardman, 2011 Shires and de Jong, 2009
	Rail	Arbantes and Wardman, 2011	
	Car	Arbantes and Wardman, 2011	
Distance		Arbantes and Wardman, 2011 Shires and de Jong, 2009 Gunn, 2001	
Data collection method	*Revealed preference*	Arbantes and Wardman, 2011 Brownstone and Small, 2005	
	Stated preference		Shires and de Jong, 2009
Income/ GDP per capita		Arbantes and Wardman, 2011 Shires and de Jong, 2009 Gunn, 2001	
Trip attributes	*Congestion*	Arbantes and Wardman, 2011	
	Transit delays	Arbantes and Wardman, 2011	
	Transit headway		Arbantes and Wardman, 2011
	Tolls		Arbantes and Wardman, 2011
Locale	*Non-Europe*		Shires and de Jong, 2009
Activities while traveling	*Productive*		Gunn, 2001 Hensher and Wang, 2016

3.3.3 Travel Behavior of Millennials

As millennials are coming of age and gaining a larger share of total travel in the U.S., researchers have become increasingly interested in the travel behavior of this cohort. In particular, researchers have been looking for factors that can account for the (at least

[11] Our estimations (Chapter 3.6) fall near the lower end of this range at about 19% of the hourly wage, which is crudely inferred from the annual household income.

temporary) decline in per capita car travel observed in many developed nations since the beginning of the new millennium (but with a sharp rebound observed at the beginning of 2016, with new record highs in VMT, if not yet VMT per capita, in the US (FHWA, 2016)). One controversial study (Bastian et al., 2016, 2017; Wadud and Baierl, 2017) suggests that economic factors such as fuel prices and gross domestic product can adequately explain the observed trend changes. In several other studies, however, millennials have been attributed an important role in explaining the reduced travel volumes through their decreased and delayed driving licensure (Blumenberg et al., 2012; Sivak and Schoettle, 2011; Kuhnimhof et al., 2012; Delbosc and Currie, 2013); shifts to non-automobile modes (Kuhnimhof et al., 2012); economic hardships, especially connected to employment (Blumenberg et al., 2012; Kuhnimhof et al., 2012); increased preferences for urban living (Blumenberg et al., 2012; Kuhnimhof et al., 2012); changes in social norms (Hopkins and Stephenson, 2015); and adoption of ICTs, whether for travel substitution (Sivak and Schoettle, 2011;), for travel inducement (Blumenberg et al., 2012; Hopkins and Stephenson, 2015), or, even, as new vanity/status objects (Tully, 2011).

Blumenberg et al. (2012), by analyzing National Household Travel Survey (NHTS) data via a series of binary logit models, revealed that driving alone to work was positively associated with higher income, while taking transit was more prominent among the "boomerang" youth (i.e., those who returned to live with their parents). Interestingly, their study shows that non-income related socio-economic characteristics, while having an effect on non-millennial commute mode choice, had none on young adults' choices. McDonald (2015) also used NHTS data to show that millennials' decrease in VMT comes from fewer automobile trips, rather than the shift to alternative modes. In other studies, Shannon et al. (2006), Kerr et al. (2010) and Zhou (2012) used convenience samples of university students to assess, mostly descriptively, millennials' reasons for choosing a particular mode and flexibility towards changing their choice.

To the best of our knowledge, there have been no prior studies that link together the willingness to pay for travel time and ICT-induced mode choices among millennial vs. non-millennial commuters. Synthesizing key findings from the three streams of literature reviewed, however, we make the following informed speculation: (1) since the productive use of travel time has been found to lower VOTT, (2) since ICT enables a broader spectrum of ways to use travel time productively, and (3) since millennials are likely to be more inclined than their elders to use ICT while traveling, millennials will have a lower VOTT than older commuters.

3.4 Sample Description: A Generational Portrait of Millennials

The empirical analysis of this paper is based on data collected from a survey administered in 2011-2012 in Northern California (Neufeld and Mokhtarian, 2012). The survey consisted of nine sections that canvassed such topics as general lifestyle and transportation opinions, personality characteristics, multitasking preferences, time use and waiting attitudes, travel mode perceptions, commuting and travel multitasking behavior, and socio-economic attributes. Paper questionnaires and invitations to take an online equivalent were distributed in transit vehicles and at transit stops, placed under windshield wipers of vehicles parked in carpool-reserved spots, sent to a large commute alternatives email list and other lists, sent to the members of a paid opinion panel, and mailed to a random selection of residential addresses. This approach enabled us to collect data from various segments of the population and different geographies, and in particular allowed for adequate numerical representation of less-often chosen modes. Specifically, we purposefully oversampled modes other than driving alone. Only respondents that were 18 years old and above, who commuted to work or school at least once a month, were invited to participate in the study. The final sample size for this study is 2216. Selected socio-economic descriptive statistics are presented in Table 3.2. The sample was weighted to approximately represent population commute mode shares; the model and subsequent results are based on the weighted sample.

As Table 3.2 shows, socio-economic descriptive statistics for the weighted sample are rather similar to the corresponding figures for the unweighted data. In other words, the trends in the sample are not heavily affected by weighting.

Table 3.2 – Selected socio-economic characteristics of the unweighted and weighted sample, distinguishing millennial and non-millennial segments

Characteristic	Unweighted dataset				Weighted dataset[a]			
	Millennials (496)		Non-Millennials (1720)		Millennials (525)		Non-Millennials (1691)	
Gender (N = 2196)								
Female	315	64.3%	1045	61.3%	351	67.6%	1054	62.8%
Ethnicity[b] (N = 2216)								
White	309	62.3%	1157	67.3%	315	60.0%	1112	65.8%
Black	13	2.6%	69	4.0%	17	3.2%	61	3.6%
Asian	98	19.8%	232	13.5%	113	21.5%	238	14.1%
Hispanic	51	10.3%	121	7.0%	52	9.9%	111	6.6%
Education level (N = 2216)								
High school	11	2.2%	55	3.2%	11	2.0%	63	3.8%
College	273	55.0%	949	55.2%	319	58.3%	968	58.0%
Graduate work	212	42.8%	716	41.6%	174	37.6%	682	38.9%
Occupation (N = 2209)								
Professional	181	36.5%	927	53.9%	195	37.1%	868	51.3%
Student	154	31.1%	34	2.0%	134	25.5%	27	1.6%
Manager	48	9.7%	325	18.9%	60	11.4%	338	20.0%
Sales	27	5.4%	52	3.0%	29	5.5%	64	3.8%
Service	11	2.2%	40	2.3%	16	3.1%	48	2.8%
Clerical	62	12.5%	276	16.1%	75	14.3%	273	16.1%
Other	10	2.0%	56	3.3%	14	2.7%	69	4.1%
Annual HH income (N = 2132)								
Less than $25,000	85	17.8%	42	2.5%	62	12.3%	47	2.9%
$25,000 to $49,999	105	21.9%	207	12.5%	135	26.7%	215	13.2%
$50,000 to $74,999	93	19.4%	342	20.7%	96	19.0%	340	20.9%
$75,000 to $99,999	80	16.7%	333	20.2%	88	17.4%	337	20.7%
$100,000 to $124,999	61	12.7%	295	17.8%	69	13.7%	266	16.3%
$125,000 or more	55	11.5%	434	26.3%	55	10.9%	425	26.1%
Primary commute mode (N = 2216)								
Biking	84	16.9%	107	6.2%	12	2.3%	22	1.3%
Commuter rail	32	6.5%	143	8.3%	4	0.8%	12	0.7%
Transit	125	25.2%	519	30.2%	47	9.0%	133	7.9%
Shared ride	83	16.7%	269	15.6%	83	15.8%	218	12.9%
Driving alone	172	34.7%	682	39.7%	379	72.2%	1306	77.2%
Population commute mode shares (N = 4,119,532)								
Biking	21830	2.2%	41347	1.3%				
Commuter rail	7717	0.8%	21787	0.7%				
Transit	88062	8.9%	248619	7.8%	N/A			
Shared ride	156576	15.9%	409727	12.9%				
Driving alone	710934	72.2%	2448824	77.2%				

[a] Weights were calculated based on population commute mode shares (separately for millennials and non-millennials) for 16 Northern California counties, available from the Census Transportation Planning Products, http://ctpp.transportation.org/Pages/default.aspx, ACS 2006–2010 data.
[b] Categories are not mutually exclusive.

Along with socio-economic attributes, the survey collected rich attitudinal and behavioral transportation-related data that can be used in deeper investigation of relationships and patterns within the sample. Attitudes were usually represented by answers to statements reported on an ordered scale (e.g., ranging from "Strongly disagree" to "Strongly agree") with three or five levels. Since latent constructs were purposefully tapped through multiple attitudinal statements (e.g., technological affinity can reveal itself through, among other ways, a preference for having newer IT gadgets and a desire to introduce them to friends), factor analyses were performed to uncover the higher-level attitudinal concepts. Travel multitasking was reported for the chosen primary commute mode and encompassed questions (binary variables) of what things commuters carried with them, in what activities they engaged, and what benefits and disadvantages they received from these activities. Objective mode attributes for chosen and unchosen modes, such as travel time and travel cost, were obtained through Google Maps and Bing Maps APIs by using the reported addresses (translated to XY coordinates) of residential and work locations. For more detailed discussions about factor analyses of attitudinal variables, travel multitasking behavior, and the collection of the objective mode attributes, readers can refer to Malokin et al. (2019).

Millennials in our sample also resemble those in more representative national sociological studies (e.g., Pew Research Center, 2014). They are more ethnically diverse than previous generations (see Table 3.3); however, their share of immigrants is lower. Young adults are somewhat better educated (75.8% of millennials and 70.1% of non-millennials have at least a bachelor's degree). They have a lower average household income ($73,000 and $101,000 for millennials and non-millennials, respectively), and their access to better-paying occupations is limited. The younger generation, on average, has lower bicycle ownership, and fewer of its members possess a driver's license.

Analysis of variance (ANOVA) results (Table 3.3) identified significant differences between the two segments with respect to various attitudes, personal traits and preferences. Millennials are found to be more technologically oriented and savvy, adopting novel gadgets sooner and using them at a greater scale: presence of ICT devices and their usage while traveling were more likely to be reported by younger adults[12]. Millennials, comparatively, feel more favorably toward active modes of transportation, like walking and bicycling. Counter to stereotype, however, they are less transit-oriented than older commuters.

Among personal traits, millennials are more extraverted, impatient, and perfectionistic (the last two traits load on the *Frustrated* factor). On the other hand, compared to the older generations, millennials are less organized and responsible – stereotypical traits of every emerging generation. On a surprising note, our data shows millennials to accept less risk and be less aggressive (*Risk-taker* factor) than non-millennials, perhaps influenced by coming of age during a global economic recession.

[12] It is interesting to note that *no* type of activity while traveling was *more likely to be performed by non-millennials.* Most behaviors that are commonly associated with ICT devices, and other basic activities, were significantly *more likely to be performed by millennials*: watching video, using internet, using a non-smartphone, using a smartphone, sending SMS, using a laptop/tablet, navigating with GPS, thinking / planning, playing electronic games, reading electronic materials, eating, resting, grooming, watching scenery, and daydreaming. The remaining activities measured were *performed at statistically similar rates by millennials and non-millennials*: listening to audio, talking to friends, talking to strangers, navigating with a map, playing non-electronic games, writing paper materials, writing electronic materials, reading paper materials, and exercising.

With regard to multitasking (Table 3.4), millennials are more willing to accept audiovisual background stimuli (e.g., radio or TV) than older generations. They also think that multitasking should be practiced by other people (normative beliefs); however, when it comes to their own behavior, they prefer to concentrate on one work-related "task" at a time but accept non-specified "activity" multitasking[13].

When they travel, millennials are more likely to carry a smartphone, laptop/tablet, and MP3 player, among other things, while non-millennials are more likely to have a newspaper, magazine, simple cellphone, or tablet (Table 3.4). Young adults reported several disadvantages of travel-based multitasking more often than their elders did, which could be an indication of their experience with a wider range of activities (including ICT-enabled activities) on the go.

[13] The survey included all items associated with the two most-commonly-used polychronicity scales: the Inventory of Polychronic Values (IPV; Bluedorn et al. 1999) and the Polychronic-Monochronic Tendency Scale (PMTS; Lindquist and Kaufman-Scarborough, 2007). Items associated with the IPV tended to use the terms "task" and "work", while those associated with the PMTS tended to use the word "activity", with no purpose specified. Whether representing a semantic artifact or a genuine difference, these two groups of items tended to load on different factors in the factor analysis.

Table 3.3 – Significant variations in socio-economic attributes and attitudes between weighted millennial and non-millennial segments

Variable	Variable type	Millennials (N=525)	Non-millennials (N=1691)	F/χ^2 statistic[a]	Signif.
Ethnicity					
White	Binary	0.60	0.66	6.035	0.014 **
South Asian	Binary	0.02	0.01	5.069	0.024 **
East Asian	Binary	0.20	0.13	12.818	0.000***
Hispanic	Binary	0.10	0.07	6.561	0.010 **
Immigrant	Binary	0.21	0.25	4.373	0.037 **
Education	Ordinal	–	–	87.383	0.000***
Annual HH income	Ordinal	–	–	153.726	0.000***
Occupation					
Professional	Binary	0.37	0.51	32.232	0.000***
Student[b]	Binary	0.26	0.02	341.052	0.000***
Manager	Binary	0.11	0.20	19.689	0.000***
HH bicycle ownership	Count	1.58	2.13	41.552	0.000***
Has driver's license	Binary	0.98	0.99	8.476	0.004***
General attitudes (factor scores)					
Pro-technology	Continuous	0.27	–0.06	40.090	0.000***
Pro-active modes[c]	Continuous	–0.12	–0.26	8.226	0.004***
Pro-transit[c]	Continuous	–0.46	–0.34	6.828	0.009***
Time pressure – reality	Continuous	0.20	–0.04	24.448	0.000***
Time pressure – preference	Continuous	0.13	–0.05	12.354	0.000***
Personality traits (factor scores)					
Extraverted	Continuous	0.18	–0.08	26.590	0.000***
Organized	Continuous	–0.08	0.07	8.789	0.003***
Frustrated	Continuous	0.26	–0.05	39.465	0.000***
Responsible	Continuous	–0.04	0.12	10.305	0.001***
Risk-taker	Continuous	–0.17	0.03	16.382	0.000***
Leader	Continuous	0.35	–0.09	83.088	0.000***
Like to move fast[d]	Ordinal	0.20	–0.09	35.558	0.000***
Multitasking preferences (factor scores)					
Background noise multitasking[c]	Continuous	0.25	0.04	18.719	0.000***
"Activity" multitasking[c]	Continuous	0.16	0.04	5.632	0.018 **
Normative multitasking	Continuous	0.12	–0.03	21.085	0.000***
(Work) "task" monotasking	Continuous	0.11	–0.09	17.017	0.000***

***, ** = significant at 1%, 5%.
[a] For binary, ordinal, and standardized variables, χ^2 statistic was used; for continuous variables, F-statistic was used.
[b] Only working students are included in the sample.
[c] Ordinarily, the sample mean of standardized factor scores would equal zero, and thus the means of each subsample (millennials and non-millennials) would have opposite signs (if different from zero). In the present instance, the sample mean may differ from zero for two reasons: (1) the current sample is a subset of the sample that was used for the factor analyses (N = 2849), and (2) weighting could substantially alter the contribution of each observation.
[d] Standardized single item.

Table 3.4 – Significant variations in travel-multitasking characteristics between weighted millennial and non-millennial segments

Variable	Variable type	Millennials (N=525)	Non-Millennials (N=1691)	F/χ^2 statistic[a]	Sign.
Activities while traveling					
Watching video[a]	Binary	0.06	0.04	7.196	0.007***
Using internet[a]	Binary	0.28	0.13	61.525	0.000***
Talking on phone[a]	Binary	0.37	0.30	8.937	0.003***
Using smartphone[a]	Binary	0.48	0.26	89.108	0.000***
Texting[a]	Binary	0.42	0.20	106.858	0.000***
Using a laptop/tablet[a]	Binary	0.10	0.05	15.571	0.000***
Navigating with GPS[a]	Binary	0.18	0.10	21.737	0.000***
Thinking/ planning[a]	Binary	0.83	0.74	19.677	0.000***
Gaming electronically[a]	Binary	0.09	0.05	14.889	0.000***
Reading electronically[a]	Binary	0.16	0.10	15.474	0.000***
Eating	Binary	0.52	0.43	12.034	0.001***
Resting	Binary	0.11	0.08	4.290	0.038 **
Grooming	Binary	0.10	0.06	10.299	0.001***
Watching scenery/ people	Binary	0.54	0.49	4.030	0.045 **
Daydreaming	Binary	0.54	0.42	22.866	0.000***
Carrying items while traveling					
Food	Binary	0.72	0.65	6.942	0.008***
Newspaper	Binary	0.03	0.10	27.939	0.000***
Magazine	Binary	0.07	0.11	8.900	0.003***
Laptop	Binary	0.26	0.17	22.017	0.000***
Smartphone	Binary	0.71	0.55	46.320	0.000***
"Simple" cell phone	Binary	0.21	0.31	19.182	0.000***
Electronic games	Binary	0.09	0.04	22.314	0.000***
Internet-enabled MP3 player (e.g., iPod®)	Binary	0.09	0.05	12.152	0.000***
"Simple" MP3 player	Binary	0.18	0.09	28.602	0.000***
Internet-enabled tablet (e.g., iPad®)	Binary	0.03	0.05	4.230	0.040 **
GPS unit	Binary	0.24	0.18	7.012	0.008***
No items	Binary	0.02	0.06	12.649	0.000***
Benefits of travel multitasking					
Makes unpleasant trip tolerable	Binary	0.21	0.13	16.992	0.000***
Disadvantages of travel multitasking					
No disadvantages	Binary	0.45	0.63	57.212	0.000***
Diminishes enjoyment of activities	Binary	0.05	0.02	17.581	0.000***
Creates unsafe distraction	Binary	0.17	0.12	9.867	0.002***
Fragments attention	Binary	0.11	0.08	7.518	0.006***
Takes time away from other things	Binary	0.18	0.09	29.762	0.000***
Can't perform activities as well	Binary	0.13	0.10	4.488	0.034 **
To what extent is commute favorable for travel multitasking[b]	Ordinal	2.45	2.36	26.497	0.000***

***, ** = significant at 1%, 5%.

[a] Originally, the activity was reported separately for two purposes: work and leisure/personal. For this analysis the purposes were combined.

[b] The variable ranges from "Hardly at all" (=1) to "Almost completely" (=5).

3.5 Mode Choice Model Estimation and Analysis

In this study, we model the choice of primary commute mode. Respondents reported their choices from a set of five alternatives: bicycle, commuter rail, transit (including local bus, express bus, light rail, and metro rail), shared ride (carpool, vanpool, and employer shuttle), and drive alone. The choice set composition is individual and contains two to four alternatives (the upper limit is due to having only four sets of mode perceptions available in each questionnaire); thus, the estimation operates with unequal choice sets.

The explanatory variables included in the model comprise mode-specific objective attributes (in-vehicle and out-of-vehicle travel time, and travel cost), socioeconomic characteristics (gender, license possession, and ethnicity), individuals' attitudes (mode perceptions, general attitudes, polychronicity), and the mode-specific propensity to use a laptop. The latter variable was computed as follows. For each mode-activity combination, we formulated a binary logit model using travelers' mode-specific involvement in each activity (=1 if reported, =0 otherwise), as the dependent variable. Individual characteristics such as socio-economic attributes, multitasking preferences, general attitudes and personality traits, time use expectations and preferences, and attitudes toward waiting were used as explanatory variables. Each model was calibrated on respondents who chose that mode, and the result was applied to predict the probability of performing that activity on that mode for all respondents, if they were to choose that mode (for additional information, see Berliner et al., 2015 and Malokin et al., 2019). That predicted probability is the propensity to use a laptop on that mode, which we can view as a lower bound on the desire to be productive while traveling.

There are several ways to compare the effects of travel multitasking on mode choice across different population segments. One simple way is to estimate a mode choice (e.g., multinomial logit) model separately for each segment. However, the estimated vector of coefficients, $\boldsymbol{\beta}$, is in fact $\mu\boldsymbol{\beta}$, where μ is an unidentifiable scale parameter that is associated with the assumed extreme value distribution for the error terms of the model, and is inversely related to the variance of that distribution. When only a single sample is involved, the μ can be fixed at 1 for convenience (Ben-Akiva and Lerman, 1985). When estimating separate models for multiple segments, however, it is conceivable that the underlying scale parameters would not be equal for different segments, implying heteroscedasticity (unequal error variances) across the segments. This could lead to erroneous conclusions if coefficients are compared at their face value. However, the comparison is valid if coefficient *ratios* are used (as in the VOTT formula), because the scale parameter cancels out in the ratio.

A more elegant way of dealing with segmented data is to estimate a single but fully-segmented model on the pooled sample. In this case, all parameters of the segmented model are estimated simultaneously; the use of full information for all segments yields efficient estimators; and the estimated coefficients can be compared directly. The artificial nested logit model (ANL) (Hensher and Bradley, 1993) offers the framework for this simultaneous estimation. ANL solves a hypothetical multidimensional choice problem, in which an actor makes a decision on which alternative to pick (e.g., mode) and what segment to be associated with (predetermined by the segmentation rule). The ANL tree structure consists of a number of nests (one for each segment) at the same level under the root. Each nest comprises the set of alternatives belonging to the segment. Accordingly, each observation belongs to only one nest and has a null choice set for alternatives associated with the other nests. For identification purposes, the inclusive value parameter of one nest must be set to 1 (the reference segment). Thus, in this tree structure, a freely-estimated inclusive value (IV) parameter indicates the value of the scaling parameter μ for that segment, *relative to that of the reference segment*. Unlike the case for the ordinary nested logit model, it is legitimate for an IV parameter of an ANL model to exceed unity, indicating only that the scaling parameter, and thence the variance of the unobserved influences on utility, for that segment are larger than the corresponding quantities for the reference segment (Hensher and Bradley, 1993).

To examine the differences in travel behavior between millennials and non-millennials, we used ANL and segmented MNL approaches. For ANL estimation, we expanded the choice set allowing for the joint decision of mode (5 alternatives) and belonging to the millennial cohort or not (2 alternatives). The inclusive value parameter was set to one for the non-millennial nest. The estimated inclusive value parameter was not statistically different from one, supporting the hypothesis that the scaling parameter μ is the same across the two segments. Accordingly, the ANL estimation degenerated into two segmented MNL models, estimated simultaneously. Although not a foregone conclusion, this result is not surprising, given that the data was collected using the same methods across the same region during the same period of time. Thus, we can use the simpler approach of segmented MNL estimation, knowing that the results are statistically equivalent to the full-information ANL approach.[14]

[14] We also implemented the approach of Swait and Louviere (1993), plotting the coefficients of one segment against those of the other segment. The slope of the best-fit line through those points in the *x-y* plane is an

For the segmented MNL estimation, we divided the sample into two parts (millennials and non-millennials) based on the year of birth of the respondents. Then, we estimated MNL models for each segment and for the whole sample. Although we tested more sophisticated model forms (including nested, cross-nested, and generalized nested logit; mixed logit; and latent class models), none of them proved statistically superior to the MNL model, perhaps an indication of good specification of the latter (see, e.g., Train's admonition (2009, pp. 35-36) that "In a deep sense, the ultimate goal of the researcher is to represent utility so well that the only remaining aspects constitute simply white noise; that is, the goal is to specify utility well enough that a [multinomial] logit model [rather than a more complex specification] is appropriate").

The three final specifications were kept identical to facilitate comparisons. Each model was weighted using the population mode shares, as described in Section 3, to account for (the purposeful) sampling bias, i.e., underrepresenting driving alone commuters and overrepresenting transit users.

All coefficients in the mode choice models have the expected sign and are significant in the pooled model. Consistency with the Independence of Irrelevant Alternatives (IIA) assumption was investigated by conducting Hausman-McFadden tests, and by evaluating a number of alternative model structures as indicated above. All of these tests failed to reject the null hypothesis that IIA holds in this case.

---

estimate of the factor by which the scales differ, and in our case the slope was not statistically different from 1.

Table 3.5 summarizes the results from the model estimation for the millennial and non-millennial segments and the pooled sample. For a full description of the variables included in the model, refer to Malokin et al. (2019).

The segmented models for millennials and non-millennials fit the data slightly better than the pooled one[15] but the goodness of fit for all three models is considered strong. While some of the coefficients are not statistically significant in the model for millennials, all coefficients of interest (travel time, travel cost and propensity to use a laptop) are significant, except (interestingly) for in-vehicle travel time. The magnitude of the millennials' IVTT coefficient is actually larger than those of the older commuters and of the pooled sample, so it should not be considered unimportant. Rather, its insignificance is arguably due to its larger standard error, which, in turn, is a function of the substantially smaller size of the millennials sample compared to those of the other two models. It may also reflect greater heterogeneity in the impact of IVTT on utility among millennials than among non-millennials, leading to greater uncertainty in the estimate of the single "average" coefficient for the younger group. In any case, since the ANL results showed that the scaling parameter did not differ across the segments, we compare the coefficients for the millennial and non-millennial subsamples directly.

---

[15] We compare segments by the *non*-adjusted ρ^2. The adjusted ρ^2 for the millennial segment is lower than for the pooled model because insignificant coefficients remained in the specification and the sample size substantially decreased, allowing the lack-of-parsimony penalty to play a greater role.

Table 3.5 – Weighted Multinomial Logit Mode Choice Models for the Millennial and Non-Millennial Segments and the Pooled Sample

Variable	Millennials		Non-Millennials		Pooled	
Objective mode attributes						
In-vehicle travel time, min	–0.017	(0.012)[a]	–0.016**	(0.007)	–0.016***	(0.006)
In-vehicle travel time (bicycle-specific), min	–0.246*	(0.130)	–0.116**	(0.047)	–0.162***	(0.060)
Out-of-vehicle travel time, min	–0.052***	(0.020)	–0.049***	(0.010)	–0.049***	(0.009)
One-way commute cost, ln($)	–1.263***	(0.307)	–1.170***	(0.155)	–1.173***	(0.138)
Socioeconomic characteristics						
Driver's license (transit-specific)	–1.205	(1.202)	–2.733**	(1.352)	–1.895**	(0.801)
Female (shared ride-specific)	–0.154	(0.312)	0.495***	(0.170)	0.362**	(0.146)
Ethnicity: white (transit-specific)	0.296	(0.439)	0.637**	(0.252)	0.561**	(0.218)
Limitation on walking (shared ride-specific)	0.205	(0.134)	0.148***	(0.058)	0.163***	(0.054)
Mode perceptions						
Mode convenience	0.431***	(0.132)	0.537***	(0.069)	0.495***	(0.059)
Mode benefit/cost	0.537***	(0.146)	0.341***	(0.078)	0.381***	(0.067)
Mode comfort	0.415***	(0.107)	0.422***	(0.065)	0.424***	(0.055)
General attitudes						
Pro-active modes (bicycle-specific)	2.842***	(0.780)	1.986***	(0.508)	2.113***	(0.406)
Pro-transit (commuter rail-specific)	1.900*	(1.013)	0.996***	(0.341)	1.138***	(0.321)
Pro-transit (transit-specific)	1.121***	(0.271)	0.775***	(0.129)	0.831***	(0.114)
Pro-transit (shared ride-specific)	0.600***	(0.171)	0.143	(0.092)	0.214***	(0.079)
Polychronicity (shared ride-specific)	0.156	(0.169)	0.215***	(0.075)	0.199***	(0.067)
Propensity for productive travel multitasking						
Propensity to use a laptop/ tablet/ netbook	2.306***	(0.558)	0.823**	(0.365)	1.238***	(0.294)
Constants[b]						
Bicycle constant	–4.914**	(1.968)	–6.218***	(1.121)	–5.411***	(1.040)
Commuter rail constant	–3.521***	(1.037)	–2.684***	(0.419)	–2.914***	(0.377)
Transit constant	0.309	(1.206)	1.683	(1.341)	0.826	(0.806)
Shared ride constant	–2.596***	(0.502)	–2.605***	(0.246)	–2.625***	(0.218)
Model information						
Number of observations	496		1720		2216	
$\mathcal{L}(\mathbf{0})$ *(varying choice sets)*	–607.885		–2033.598		–2641.483	
$\mathcal{L}(\boldsymbol{c})$ *(varying choice sets)*	–384.172		–1197.255		–1587.107	
$\mathcal{L}(\hat{\boldsymbol{\beta}})$ *w/o constants*	–289.934		–984.683		–1295.101	
$\mathcal{L}(\hat{\boldsymbol{\beta}})$	–257.137		–881.276		–1156.654	
ρ^2 *(equally-likely base)*	0.577		0.567		0.562	
Adjusted ρ^2	0.542		0.556		0.554	

Significance: *** – < 0.001, ** – < 0.01, * – <0.05.

[a] Effects of the variables are represented by an estimated coefficient and standard error (in parentheses).

[b] Driving alone is the base alternative for each model.

Owning a driver's license has a negative coefficient (with respect to the transit alternative) and is significant for non-millennials only, indicating that having a driver's license does not lower the utility of taking transit among millennials. Non-millennial women obtain higher utility than their male counterparts from taking a shared ride. The same variable has a negative but statistically insignificant coefficient for millennials, suggesting that young females derive similar utility from carpooling as do young males. An analogous situation happens with respect to ethnicity: older white adults are more likely than older non-whites to take transit (in the study area the transit network substantially – though of course not exclusively – serves affluent, predominantly white, residential areas), while the insignificant coefficient in the millennial segment suggests that for younger adults, ethnicity does not play a role in choosing between riding transit and the reference alternative of driving alone.

Considering mode perceptions, millennials are noticeably more sensitive than non-millennials (the coefficient is higher) to a mode's *benefit/cost* factor score, while non-millennials value mode convenience more than millennials do. This effect may reflect millennials' search for a mode with the best value (e.g., lower costs, greater benefits) even if it is less convenient. The *pro-active transportation* and *pro-transit* attitudes have greater positive coefficients for biking, commuter rail, transit, and shared ride alternatives for millennials, making them more likely to choose these non-drive alone modes compared to older commuters with similar attitudes.

3.6 Value of Travel Time and Willingness to Pay for Laptop Usage

Using the results in Table 3.5, we can compare the value of travel time (VOTT) and willingness to pay (WTP) for laptop usage, for the members of the two generational groups. VOTT evaluates the tradeoff, or substitution, between the time and cost of a trip, i.e., how much travelers are willing to pay (be paid) to reduce (increase) their commute time, in order to leave their utility constant. We calculate the VOTT for mode *m* as the ratio of the derivative of that mode's utility with respect to its travel time (IVTT or OVTT) to the derivative with respect to its cost (Koppelman and Bhat, 2006):

$$VoTT_{TT,m} = \frac{\partial U_m}{\partial TT_m} \Big/ \frac{\partial U_m}{\partial Cost_m} = \frac{\partial U_m}{\partial TT_m} \Big/ \left(\frac{\partial U_m}{\partial \ln Cost_m}\right)\left(\frac{\partial \ln Cost_m}{\partial Cost_m}\right) = \frac{\beta_{TT,m}}{\beta_{\ln Cost,m}} * 60\,\frac{min}{hr} * Cost_m, \tag{3.1}$$

where $VoTT_{TT}$ is the value of travel time for either in-vehicle or out-of-vehicle travel time measured in 2011 US dollars per hour, β_{TT} is the estimated coefficient for IVTT or OVTT (each measured in minutes), $\beta_{\ln Cost}$ is the estimated coefficient of the natural logarithm of travel cost, and $Cost$ is the individual-specific one-way trip cost expressed in 2011 US dollars. Since travel cost is represented by the natural logarithm of the one-way cost of the commute, the utility function is not linear in cost, and thus VOTT varies by individual: all else equal, the greater a commuter's cost for a trip of a given length, the more she is willing to pay (requires being paid) to save (incur) a fixed amount of time. Table 6 reports the mean and median of the weighted distribution of VOTT for each sample.

We use the WTP for productive multitasking to evaluate the substitution between the propensity to use laptop and either time or cost. In other words, this is a measure of how much (measured in either additional trip time or cost, respectively expressed in minutes and 2011 US dollars) commuters are willing to pay (or be paid) for the ability to use a laptop on the commute. The propensity to use a laptop, which varies between zero and one, is intrinsically composed of two parts: the general conduciveness of the mode to using a laptop and the individual's inclination to engage in this behavior. Thus, it is more realistic to assume a certain cap on the rate of substitution. For example, if allowed to reach the extreme upper value (propensity=1), WTP measures could suggest practically unachievable substitution rates for the target modes (having an absolute conduciveness to use a laptop) or individuals (assuming the maximum inclination for everyone). This is the rationale behind introducing an additional factor into the WTP calculation (Equation 3.2). This factor is the difference in the propensity to use a laptop between the *reference* mode and another, "*target"*, mode. Commuter rail is universally used as the reference mode for conceptual reasons: it presumably provides the best experience for productive multitasking among all modes that are considered. Thus, this formulation of the WTP measures how much travelers are willing to pay in terms of time or money *to obtain the same level of productive multitasking on the target mode that they would achieve on commuter rail*, given their individual tendency for this behavior. For money, the expression is:

$$WTP_{R,k}^{Cost} = \frac{\beta_{Laptop}}{\beta_{\ln Cost}} * Cost_k * (Laptop_R - Laptop_k), \tag{3.2}$$

where $WTP_{R,k}^{Cost}$ is the willingness to pay in terms of trip cost with the reference mode R and target mode k, β_{Laptop} is the estimated coefficient of the propensity to use a laptop, and $Laptop_R$ and $Laptop_k$ are the propensities to use a laptop for the reference and target modes respectively. Cost, the laptop variables, and hence the WTP, vary by individual. For time, the equation loses the "cost" factor, and $\beta_{\ln Cost}$ is replaced by β_{IVTT}.

We selected two target modes of interest: driving alone and transit. In our formulation, WTP for using a laptop while driving alone implies the adoption of fully autonomous vehicles (AVs), which allows users to experience the same level of multitasking conduciveness that is observed for commuter rail (i.e., an ability to divert attention from the driving task to the cognitively demanding tasks that using a laptop requires). In essence, $WTP_{Rail,DA}^{x}$ measures the premium (in x = minutes, dollars) that commuters would be willing to pay for productive multitasking while "driving alone" in an AV. Similarly, WTP for productive multitasking while taking transit evaluates the premium that commuters are willing to pay for productive multitasking on public transportation. This could be achievable with currently available means, such as providing more room and seating on vehicles, facilitating internet connectivity with accessible Wi-Fi and electric outlets, etc.

In Table 3.6 we present estimated weighted VOTT and WTP measures for the millennial and non-millennial segments. The discrepancies between the mean and median values for the VOTT and WTP estimates (the mean always being larger), which are more prominent for the latter, arise from highly skewed distributions with heavy right-hand tails. These distributions point to a considerable level of heterogeneity in both segments. Existence of negative WTP values for a sizable number of individuals (which occurs when $Laptop_R < Laptop_k$), indicates that for these individuals, the target mode is already more conducive to productive multitasking (in terms of their specific proclivities) than commuter rail is, and they would need to be paid in order to accept a target-mode conduciveness that is equivalent to that of the inferior (on this dimension, for them) commuter rail value. These results point to the importance of examining variability, not just means, and provide a strong indication that individuals do not always conform to our stereotypical expectations. For more distributional statistics, see Tables Table B.1, Table B.2, and Table B.3 in Appendix B.

Table 3.6 – Weighted VOTT and WTP for Productive Multitasking for the Millennial and Non-Millennial Segments and the Pooled Sample

Parameter	**Millennials**		**Non-Millennials**		**Pooled**	
N	496		1720		2216	
ρ^2	0.577		0.567		0.562	
β_{IVTT}	–0.0171†		–0.0160		–0.0165	
β_{OVTT}	–0.0524		–0.0493		–0.0491	
$\beta_{ln\,Cost}$	–1.263		–1.170		–1.173	
β_{laptop}	2.306		0.823		1.238	
	Mean	Median	Mean	Median	Mean	Median
$WTP(laptop)^{\$}_{Rail,DA}$	0.55	0.06	0.30	0.02	0.42	0.02
$WTP(laptop)^{\$}_{Rail,Transit}$	0.21	–0.03	0.16	–0.03	0.21	–0.03
$WTP(laptop)^{min}_{Rail,DA}$	14.24	3.66	4.50	0.92	6.86	1.45
$WTP(laptop)^{min}_{Rail,Transit}$	3.33	–2.90	0.60	–1.60	1.07	–2.20
VOTT for IVTT, \$/hr	1.87	1.50	2.20	1.64	2.18	1.66
VOTT for OVTT, \$/hr	5.70	4.59	6.76	5.04	6.50	4.93

† Significant at <17% level.

Both cohorts, millennials and non-millennials, view their OVTT as more onerous than IVTT, which is consistent with expectations: the OVTT coefficients are more than three times greater in absolute value than the respective IVTT coefficients in each age group. However (recalling from Section 4 that the coefficients for the two segments have statistically equivalent scales and thus can be compared directly), the differences in respective coefficients *between the cohorts* are much smaller at 6.9%, 6.3%, and 7.9% (higher magnitudes for millennials) for the IVTT, OVTT, and cost variables, respectively. Only coefficients for the propensity to use a laptop are substantially different across the age groups, at 2.8 times higher for millennials. When coefficients are estimated simultaneously in the ANL model formulation, chi-square tests for the restricted versus unrestricted specifications show that IVTT, OVTT, and cost coefficients are not significantly different between the segments (the chi-squared test statistic for comparing the restricted and unrestricted models is 0.00164 if all three coefficient pairs are equated), while the coefficient for the propensity to use a laptop is. The directionality of these differences suggests that millennials are more sensitive to all four variables, as they are slightly more averse to increases in a mode's IVTT, OVTT, and cost, and considerably more responsive to increases in the propensity to use a laptop on that mode. This finding indicates that millennials are slightly more "unhappy" about the travel time and cost of a mode than the older generations are, but they are much more willing than their elders to "tolerate" that mode if they can spend their travel time productively using a laptop or tablet on it.

Interestingly, despite the slightly higher-magnitude coefficients for millennials, the mean and median VOTT for IVTT and OVTT are *lower* for millennials than for non-millennials. In other words – as conjectured at the outset – millennials, on average, are not willing to pay as much to save each minute of travel as older adults are (15.0 and 15.7% less for IVTT and OVTT, respectively). When the VOTT computation is broken down into factors, $VoTT_{IVTT,m} = \frac{-0.0171}{-1.263} * 60\frac{min}{hr} * Cost_m = 0.81235 * Cost_m$ and $VoTT_{IVTT,m} = \frac{-0.0160}{-1.170} * 60\frac{min}{hr} * Cost_m = 0.82051 * Cost_m$ for millennials and non-millennials respectively, it is apparent that the difference overwhelmingly lies within the distribution of the cost variable between the segments.

We further investigated the chosen mode costs, focusing on the drive-alone mode since it accounts for a 77% weighted share. The overall drive-alone cost is lower for millennials than for the older generations, and an even deeper investigation of the cost components indicated that millennials tend to have lower commuting costs for each component. Particularly, compared to non-millennials, millennials drive slightly (2%) more fuel-efficient vehicles, they pay slightly (3%) lower amounts to tolls, they take slightly (1%) shorter commutes, and they pay substantially (21%) less for parking. All these effects compounded cause the average weighted VOTT for millennials to be lower than that for non-millennials.

While the valuation of travel time between segments does not paint an entirely straightforward picture, willingness to pay for using a laptop does. Specifically, millennials consistently have a higher willingness to pay for using a laptop while traveling compared to the non-millennial cohort. On average, they are willing to pay for the ability to use a laptop in a(n autonomous) vehicle \$0.55 (or 14.2 minutes) per one way commute, while non-millennials are willing to pay only \$0.30 (or have a commute that is 4.5 minutes longer). Similarly, in the case of public transit, millennials are willing to pay \$0.21 (or 3.3 minutes) for the same ability, while non-millennials would make the same substitution only for \$0.16 or 0.6 minutes per commute. All these estimates show that millennials are more sensitive to the ability to use a laptop/tablet/netbook while commuting, and accordingly present a potential market for multitasking-friendly travel options (e.g., autonomous vehicles, ride-hailing alternatives, public transit improvements, etc.).

A critic might point out that neither of our models controls for income, and therefore wonder whether the difference between the market segments could be simply explained by the fact that millennials are, on average, earning less than their career-advanced counterparts. However, our experimentations showed that including attitudinal variables in the model rendered household income indicators insignificant for every segment. Moreover, stratifying the estimated VOTT and WTP measures by cohort, chosen mode, and household income categories exhibits no clear relationship among these variables. As the next section shows, the attitudinal variables themselves make a pronounced contribution to the observed discrepancies between millennials and non-millennials with respect to VOTT and WTP measures.

3.7 Sensitivity Analysis

Our estimated VOTT measures (Table 3.6) are on the lower side of the range found in the literature. On the one hand, if meta-analysis models are applied, the expected VOTT of IVTT would be \$3.78/hr and \$4.81/hr (Shires and de Jong, 2009 and Arbantes and Wardman, 2011, respectively), correcting for historical currency exchange rates and inflation – values that are not out of scale with ours. On the other hand, our preferred MNL model, which contains attitudinal and multitasking attributes (Table 3.5), produces even lower VOTT measures (Table 3.6) than the meta-analysis would suggest. These additional variables, which are absent from most specifications found in the literature and practice, could decrease VOTT estimates by interacting with travel time and travel cost and decreasing their direct effects on utility.

To evaluate the effect on the VOTT estimates of having different combinations of the explanatory variables, we tested several specifications of the MNL model. In Table 3.7, the first specification contains only mode attributes (travel time and travel cost) and socio-economic characteristics as explanatory variables, representing a conventional formulation of the model. The VOTT of IVTT for the pooled model, $3.36/hr, is very close to the meta-analysis estimates mentioned above. There is a substantial difference between the VOTT of IVTT for the millennial and non-millennial segments, with the former (counterintuitively) being 69% *higher* than the latter. Compared to this divergence, VOTT measures associated with OVTT are virtually equal between the segments.

A much greater impact is associated with the inclusion of the propensity to use a laptop along with the mode attributes and socio-economic characteristics in the model (Specification 2). While non-millennials do not demonstrate any large shifts in either VOTT estimate, the VOTT with respect to IVTT for the millennial group decreases by $1.29/hr (26%) and the VOTT for OVTT increases by $0.52/hr (10%).

Finally, for completeness, the third specification includes all explanatory variables from Table 3.5 and replicates the VOTT and WTP ratios from Table 3.6. By including attitudinal variables (mode perceptions and general attitudes) in the pool of explanatory variables, the mean VOTT with respect to IVTT for the millennial segment decreases substantially (by $1.85/hr, or 50%), with a coincidentally commensurate increase (by $1.46/hr, or 28%) in the mean VOTT with respect to OVTT for the non-millennial segment.

In this model alone, the VOTT for IVTT and OVTT are *lower* for millennials (by 15-16%) than for non-millennials. It should be emphasized that for *both* specifications that include the laptop variable (i.e., models 2 and 3), the coefficient of that variable is substantially (2-3 times) larger for millennials than for non-millennials, leading to a substantially higher WTP (in either time or money) to use laptop for the younger group. However, for Specification 3, the WTP for using a laptop experienced a decrease in terms of monetary valuation and an increase in temporal valuation, for both segments.

It is tempting to analyze changes in the TT, cost, and laptop coefficients themselves across the specifications of Table 3.7, but this would not be appropriate. The reason is that moving variables (attitudes and the laptop propensity) from unobserved ("ε") to observed ("V", in the common notation for discrete alternative utilities) changes $Var(\varepsilon)$ and thence the scale of the coefficients in the model (Ben-Akiva and Lerman, 1985) – possibly differently for each segment. Thus, even within a given segment, when we see a coefficient decrease or increase across specifications, we are seeing the confounding of scale changes and true changes in coefficients as a function of correlations among the included (versus excluded) explanatory variables. When comparing *across* segments, we have the added confound that perhaps the scale is changing differently across segments as we change the specification. We could estimate an ANL model on each specification, but that would only tell us whether the scales continue to be equal across segments (and if not, the size of the scale of one segment *relative to that of the other*), not whether the scales are increasing or decreasing *in absolute terms* as we include more variables. By confining our attention to changes in VOTT and WTP, we avoid this issue, since the unknown scale parameter cancels out in the ratio of two coefficients.

With that in mind, considering the range of specifications shown in Table 3.7, we can summarize the trends as follows: failing to separate out the effects of multitasking-related variables overestimates both cohorts' VOTT for IVTT and underestimates it for OVTT. The overestimation of in-vehicle VOTT is substantially greater for millennials (by a factor of more than 2.5, between specifications 1 and 3) than for non-millennials (by 35%, between the same two models), while the underestimation of out-of-vehicle VOTT is substantially lower for both segments, but somewhat greater (23%) for non-millennials than for millennials (12%).

Table 3.7 – Sensitivity Analysis of Weighted Mean VOTT and WTP for Laptop Usage for Millennial and Non-Millennial Segments and Pooled Sample

Speci-fi-cation	Included variables	Segments	Estimated ratios									
			β_{IVTT}	β_{OVTT}	$\beta_{ln\,Cost}$	β_{laptop}	Value of IVTT, $/hr	Value of OVTT, $/hr	$WTP^{\$}_{Rail,DA}$	$WTP^{\$}_{Rail,Tra}$	$WTP^{min}_{Rail,DA}$	$WTP^{min}_{Rail,Tra}$
1	*Socio-economic Objective mode attributes*	*Mill*	–0.0364***	–0.0365**	–0.998***	–	5.01	5.02	NA	NA	NA	NA
		Non-mill	–0.0244***	–0.0429***	–1.324***	–	2.96	5.19	NA	NA	NA	NA
		Pooled	–0.0270***	–0.0414***	–1.248***	–	3.36	5.15	NA	NA	NA	NA
3	*Socio-economic Objective mode attributes Laptop usage propensity*	*Mill*	–0.0286***	–0.0426**	–1.058***	2.681***	3.72	5.54	0.76	0.29	9.94	2.33
		Non-mill	–0.0232***	–0.0436***	–1.319***	1.142***	2.81	5.30	0.37	0.20	4.33	0.58
		Pooled	–0.0247***	–0.0430***	–1.255***	1.572***	3.05	5.32	0.50	0.25	5.82	0.91
3 (final model of Table 3.6)	*Socio-economic Mode attributes Laptop usage propensity Attitudes*	*Mill*	–0.0171	–0.0524***	–1.263***	2.306***	1.87	5.70	0.55	0.21	14.24	3.33
		Non-mill	–0.0160**	–0.0493***	–1.170***	0.823**	2.20	6.76	0.30	0.16	4.50	0.60
		Pooled	–0.0165***	–0.0491***	–1.173***	1.238***	2.18	6.50	0.42	0.21	6.86	1.07

Significance: *** – < 0.001, ** – < 0.01.

3.8 Discussion and Conclusions

The millennial generation is a target population for many ICT technologies that have gained prominence within the last decades. As these new ways to create, transfer, and consume information streams permeate modern lifestyles, it is inevitable to observe some effects on consumed resources (e.g., time, attention) and gained benefits (e.g., productivity, happiness) with respect to travel behavior in general and mode choice in particular. Specifically, travel-based ICT use, or multitasking, could be expected to modify the influence on mode utility of objective travel characteristics, especially to diminish the monetized valuation of travel time – an effect which has long been conjectured in the literature.

Using a revealed preference commute mode choice MNL model estimated on data collected from a specially-designed survey, we investigated the intersection of these three timely travel behavior topics: the impact of activities while traveling on mode choice, the estimation of WTP and VOTT, and the analysis of the travel behavior of millennials – to our knowledge, the first empirical study to do so. Our comparison of millennial and non-millennial commuters in Northern California found that the younger adults show shifting tastes with respect to influences on their travel behavior, with important implications for planning and modeling purposes.

First, in stark contrast to convention and to the older commuters in our sample (for whom driver's license possession, gender, ethnicity, and walking limitations were significant), we could not find any significant influence of socio-economic variables on the mode choice of younger adults. Second, with respect to subjective perceptions of commute alternatives, millennials are notably less sensitive to convenience and more sensitive to perceived benefits/costs than their older counterparts are. Third, a given level of support for active transportation and for transit has a stronger (positive) influence on the utilities of those modes for millennials than for non-millennials. Fourth, a given propensity to use a laptop or tablet on a certain mode has a much stronger (positive) influence on the utility of that mode for millennials, with the result that they are willing to pay more in time or money than their older counterparts for the ability to use a laptop/tablet while commuting. Finally, VOTT averages for both in-vehicle and out-of-vehicle travel time are lower for millennials than for older adults. Unfortunately, current state-of-practice travel behavior models currently lack many of these variables, including the propensity for travel-based multitasking among other subjective characteristics.

The lower VOTT observed for millennials is consistent with expectations, but further investigation found it to be lower for reasons which, at first glance, have nothing to do with travel-based multitasking. Specifically, it is lower because (due to our log transformation of cost in the utility equations) it is a function of cost, and commuting costs tend to be lower for the millennials in our sample. Since this may be partly a function of their junior status in the workforce, the effect is not likely to persist over time. Looking more closely, however, we realize that a differential effect of travel multitasking is manifested in at least two ways – one directly, and one more indirectly. First, as mentioned above, the willingness to pay (either in time or in money) for the ability to use a laptop on the commute is markedly higher for millennials than for older commuters. Secondly, a sensitivity analysis showed that when multitasking perceptions, preferences, and propensities are *not* separately accounted for (i.e., when those effects are absorbed into the coefficients of travel time and cost in particular), counterintuitive results emerge.

Specifically, the IVTT VOTT for millennials is (1) *higher* than that of non-millennials, and (2) equal to, instead of lower than, that of millennials' *OVTT* VOTT. This highlights the value of accounting, and the need to account, for these benefits in an explicit way, as our final model does.

The findings of this study portend diverse shifts in future travel demand, potentially including (1) proliferation of urban sprawl due to an increased willingness of travelers to accept longer and costlier commutes; (2) increased market shares of modes that are suitable for travel-based multitasking (public transit nowadays, autonomous vehicles in the future); and (3) induced travel demand due to the ever-increasing volume of travel-related information (e.g., the availability of attractive destinations).

Several limitations of this study could affect the generalizability of the results. First, the data on which the analysis is based is obtained from a relatively affluent and urbanized slice of the U.S. population. The available transportation options and their characteristics, lifestyles and employment, access to ICTs, and the usage patterns of the latter will differ across the United States and globally. Second, as information technology advances rapidly, consumer ICT products and services have a very short lifecycle[16]. Taking this into account, it is reasonable to wonder whether data collected in 2011-2012 produces results that are relevant today. We argue that while the technology changes rapidly, the underlying purposes for which this technology is being used (e.g., productive, recreational, etc.) are much more stable, which allows transferability of the results over time. We also argue that

[16] While the brand or form factor (a hardware design standard that shares similar size, shape, and other physical specifications, e.g., a desktop tower, flip phone, etc.) of a certain ICT product could be the same across a number of years, the content and the scope of such a product would, most likely, be quite different. For example, compare the heavy and bulky laptops from the early 1990s to the ultra-thin and light laptops of the late 2010s; or compare the range of functionality of early smartphones from the mid-2000s to the latest consumer offerings today.

the *methodology* itself has persistent value, to the present type of application as well as to many others. Third, even though we collected observations about more than 20 activities commuters could perform while traveling, we focus only on one (using a laptop/ tablet/ netbook) to analyze the impact on VOTT of travel-based multitasking. This choice was guided by the multicollinear relationships between the activities (e.g., various types of similar activities performed with the same medium) and by targeting *productive* travel-based multitasking, as it, arguably, may have the most prominent impact on mode choice and VOTT.

To address these limitations to the current study, future research would benefit from broadening the scope by evaluating the impacts of other forms of travel-based multitasking on mode choice and VOTT. Methodologically similar investigations in the U.S and around the world could uncover heterogeneity across various populations and geographies. Additionally, these investigations would allow understanding historical trends of activity engagement across multiple generations (i.e., including the latest newcomer – Generation Z).

3.9 Acknowledgements

The funding for this project was provided by the National Center for Sustainable Transportation at UC Davis (which receives funding from the U.S. Department of Transportation and the California Department of Transportation through the University Transportation Centers program); a new faculty grant from the Georgia Institute of Technology; and the Capitol Corridor Joint Powers Authority (CCJPA). We are grateful to Amanda Neufeld and Zhi Dong, who were deeply involved in the survey design, and organized the data collection and cleaning effort. We wish to thank Rosaria M. Berliner and Judith A. Cruz, who contributed to the data modeling, and Javier R. Salgado, who helped with descriptive analysis and additional data cleaning.

CHAPTER 4. TRANSFER LEARNING FOR ENRICHING NHTS WITH ATTITUDINAL DATA

Malokin, Aliaksandr, Patricia L. Mokhtarian and Giovanni Circella (2018b) An investigation of methods to enrich National Household Travel Survey data with attitudinal variables. Available from the authors.

4.1 Abstract

Often in practice, the problem of unavailability of specific desired knowledge within one ("target") dataset arises. However, if this knowledge can be extracted from a different ("source") dataset and transferred between the datasets, this could increase the value of the target dataset at relatively minimal cost. The goal of this paper is to evaluate approaches to informing one dataset with knowledge from another and to evaluate the performance of the knowledge transferred into the target dataset. We use the 2009 National Household Travel Survey as the target dataset. The missing knowledge is transportation-related attitudes, whose inclusion could greatly improve travel behavior models. Our source dataset is obtained from the 2011–12 Multitasking Survey of Northern California Commuters. To achieve the goal, the set of common variables was first augmented with a large number of built-environment attributes. Then, after applying machine-learning methods, *pro-transit*, *pro-active transportation*, and *pro-density* attitudinal factor scores were predicted with the greatest precision; correlations of the predicted and observed scores were 0.564, 0.538, and 0.571, respectively. The performance of the transferred attitudes was measured by estimating linear regression models of vehicle ownership. The

results showed that in the source dataset the observed attitudes account for an 8.0% model lift (improvement in goodness of fit), while in the target dataset the predicted attitudes account for a 1.2–5.4% model lift. Although these initial results are modest, we believe they show substantial promise, and the process has identified a number of opportunities for improvement and further research.

4.2 Introduction

Travel demand forecasting and travel behavior modeling experience both the benefits and disadvantages associated with the increased data availability of the information age. Embracing new data acquisition techniques, such as GPS-based trajectory records of movement, smartphone geolocation, Bluetooth, and Near Field Communication sensing, has been a pioneering effort that allows gathering more travel behavior data while keeping the respondents' burden at a minimum. However, many important factors that influence where and how people travel lie outside of manifest travel behavior dimensions, and are still mainly collected in the form of self-reported, disaggregate survey data. Among these factors, we consider lifestyles, attitudes, motivations, intentions, and similar constructs to be especially critical.

Despite the development of internet-based surveys and smartphone-based lightning polls, a crucial problem with this type of data still exists: there is a direct relationship between the amount of useful information to be collected from respondents and their resource burden during this process, and correspondingly an inverse relationship between that burden and the likelihood of obtaining the desired information. For decades, a quest for the optimal balance, given fixed (and modest) budgets, forced investigators to target narrower topics and sacrifice breadth for depth (or, vice versa) with respect to the collected information. For example, the 2009 National Household Travel Survey (NHTS), which surveyed more than 150,000 households in all 50 U.S. states, collected mainly socio-economic characteristics and observed travel behavior attributes. Alternatively, numerous researchers collect much smaller samples, generally within a limited geographical area, studying travel behavior phenomena and measuring numerous explanatory variables that are not captured by the NHTS.

In this study, we implement and evaluate a number of methods for using a sample containing attitudinal measures among other variables (the "source dataset"), to predict attitudes for the observations in an unrelated dataset (the "target dataset"). The choice of the NHTS as the target for the transferred information was motivated by its importance to many transportation studies in the United States and its value to the agency funding this work. The attitudinal data source is the travel-multitasking survey administered by the authors in Northern California in 2011-2012 (referred to as the Multitasking Survey of Northern California Commuters – MSNCC, in the remainder of the paper).

To inform one dataset (NHTS) with the information available in another (MSNCC) and evaluate the performance of this process, we propose two separate frameworks. The first one is the *transfer learning framework*. It is tasked to robustly evaluate the performance of predicting functions given the knowledge to be transferred (attitudes) and the pool of common variables (socio-economic and land use). The second one is the *external validation framework*. In the context of the target dataset, it assesses how valuable the transferred knowledge is for model building. These frameworks are developed to be readily transferrable beyond the context of this study and can be applied in various settings where one dataset is merged with the variables from another via statistical inferences.

The rest of this paper is organized as follows: Section 4.3 formally defines transfer learning and provides some background on statistical matching, data fusion, and key machine learning concepts and on the methods that are used in this study. Section 4.4 identifies the working substrate of this study (the NHTS and MSNCC datasets) and establishes the transfer learning and external validation frameworks, which are responsible for enriching the NHTS dataset with attitudes and evaluating their performance, respectively. Practical details of applying the transfer learning framework and the subsequent results are laid out in Section 4.5. In Section 4.6, we describe and discuss the results of the external validation exercise, implemented as a vehicle ownership model. Section 4.7 summarizes the results of the study and highlights avenues for further research. In Appendix C we provide expanded detail on the literature review (offering something of a mini-tutorial on transfer learning methods, for readers who may be unfamiliar with them), our transfer learning application for categorical variables, and supporting tables.

4.3 Brief Background and Review of Related Literature

4.3.1 An Overview of Approaches to Combining Datasets

For several decades, there has been an interest in combining independently-collected datasets and providing a “one stop shop” for a specific set of data needs. This interest only flourished as data-derived insights became more attainable and expected for decision making. The germinal attempts at data matching in the 1960s coincided with the initial spread of accessible computing power capable of handling big datasets (hundreds of thousands of records). Pioneered by governmental organizations (in the U.S., the Social Security Administration and the Internal Revenue Service) – the original "big data" powerhouses – these initial studies aimed to bridge tax, income, and demographic records for the purposes of filling in missing information, finding discrepancies in the reported data, and tracking taxpayers over time to investigate longitudinal trends. Interestingly, as Okner (1974) reported, there was a sentiment of doubt among researchers as to whether obtaining synthetic data through matching was any better than direct surveying, given the substantial amount of human and computational resources enlisted by the former.

Since these fledgling inquiries, the terminology behind the concept of informing one dataset with another has been multifarious, with several contenders coined by different scientific groups. Early into the research, *record linkage*, or *exact matching*, or *exact linking* (of records that describe the same entities; Newcombe et al., 1959) was distinguished from *synthetic (stochastic) linking* or *data synthesizing* (of records that are matched via some approximation; Okner, 1974). Later, while *record linkage* gained ground and blossomed fruitfully over the years (Winkler, 1999), the term *synthetic linking* fell out of fashion in favor of *file concatenation* (e.g., Rubin, 1986), *data fusion* (e.g., Baker et al., 1989) *statistical matching* (e.g., D'Orazio et al., 2006), *ascription* (e.g., van der Putten and Kok, 2010), *data augmentation* (e.g., Hüttenrauch, 2016), and *data triangulation* (e.g., Hand, 2018). For clarity of presentation, this work will adopt *statistical matching* to serve as a "catch-all" term for this process.

At the same time, originating in military applications, the term *data fusion*, "a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats as well as their significance" (White, 1991, p. 5), got its footing in the fields of signal processing, statistical inference, and machine learning, usually as an abbreviation of the longer term *multisensor data fusion* (Hall and Llinas, 1997; Castanedo, 2013; Khaleghi et al., 2013). The digitization of the modern economy and the boom in consumer information and communication technology (ICT) devices expanded non-military applications of *multisensor data fusion* that now include, for example, lifestyle and medical trackers (Gravina et al., 2017); intelligent transportation systems (ITS) and traffic management (El Faouzi et al., 2011); and autonomous driving (Becker and Simon, 2000).

By using data from several sources that characterize identical entities, *multisensor data fusion* improves the confidence in and reliability of pattern detection, and is more akin to *record linkage*. The first mentions of *data fusion* as a synonym for *statistical matching* can be traced back to works of French and German market researchers in the late 1970s and 1980s (as referenced in Baker et al., 1989; and Rässler, 2002), while *multisensor data fusion* appeared on the radar at least as early as the mid-1980s (Waltz, 1986). Today it is unclear which field has a greater right to claim the definitive terminology, but the detrimental effects of their concurrent existence are apparent. The two originally distinct processes of *record linkage* and *statistical matching* have been conflated via an enveloping term *data fusion*, which hampered diffusion of knowledge within research communities and resulted in the proliferation of endemic studies, which are poorly aware of developments in the other fields. As Khaleghi et al. (2013, p. 28) put it: "Data fusion is a wide ranging subject and many terminologies have been used interchangeably. These terminologies and ad hoc methods in a variety of scientific, engineering, management, and many other publications, shows the fact that the same concept has been studied repeatedly."

The panoply of terms describing the process of multi-source data integration could be an indirect testament to this argument. The present authors are far from the first to be perturbed by the lack of consistency in the terminology: it has previously been pointed out by Rässler (2002), D'Orazio et al. (2006), and Tsamardinos et al. (2012), for example. To avoid proliferating confusion, this work will refrain from using *data fusion* to describe exclusively the statistical matching process, due to its broader nature and conflated usage. However, readers should be aware that the practice of equating *data fusion* and *statistical matching* is still widespread, especially in European marketing research literature (e.g., Kamakura and Wedel, 1997, van der Putten, 2002; Rässler, 2004; van der Putten and Kok, 2010; Fisseler and Feher, 2010).

Not surprisingly, the problem of fusing data has been also studied within the computer science field, which led to the development of its own distinct methodology. In keeping with the general terminological theme (compare "machine learning", "supervised learning", "deep learning", etc.), the computer-science-based methodology of bridging knowledge sources (i.e., different datasets) to improve task performance (i.e., predictive function accuracy), is fittingly labeled *transfer learning* (Pan and Yang, 2010). It borrows heavily from adaptive behaviors observed in the biological world, in which actors transfer their previously learned skills into new settings. Both "flavors" of *data fusion* (*multisensor data fusion* and *statistical matching*) could be encompassed by *transfer learning* (Zheng, 2015). Stemming from the computer science field, *transfer learning* is innately posed to implement machine learning methods that are capable of handling large amounts of information (i.e., *big data*) computationally efficiently. However, the compartmentalization of the fields is persistent: to our knowledge only three published works (Tsamardinos et al., 2012; Lagani et al., 2016 – explicitly; and Chen et al., 2015 – implicitly) have acknowledged the coexistence of *statistical matching* and *transfer learning* and applicability of the latter to statistical matching problems.

4.3.2 Transfer Learning: Definitions, Terminology, and Key Concepts

In this study we aim to implement a statistical matching application by using transfer learning. *Transfer learning* is a machine-learning framework that defines the formal means of knowledge transfer between domains (datasets, or *variable spaces*) using *tasks* – combinations of predictive learning methods and target variables.

Following the transfer learning framework outlined in Pan and Yang (2010), we begin with the concept of a *variable space* $\mathcal{X}$, which is the set of all available variables of interest to a study. A specific $n \times p$ data matrix to be analyzed is denoted X, whose *n* rows constitute *n* cases or observations on *p* variables of interest, i.e., *n* particular elements of a *p*-dimensional $\mathcal{X}$ or of a *p*-dimensional subspace of a larger-dimensional $\mathcal{X}$. The values of *n* and *p* could change in the course of the analysis, as cases are filtered out (or, less commonly, added) and variables are added or dropped (see discussions in Sections 3.1 and 4.1). A *domain* D consists of a variable space $\mathcal{X}$, and probability distribution $P(X)$ over the *n* observations of a specific X matrix to be analyzed. The simplest case of transfer learning involves two datasets. For a source domain, $D_S = \{\mathcal{X}_S, P(X_S)\}$, let the mapping between it and the variables of interest to be transferred, $\mathcal{Y}_S$, be known. Let a target domain, $D_T = \{\mathcal{X}_T, P(X_T)\}$, contain the other dataset, which will be the recipient of the transferred information. Then, if there is an intersection between D_S and D_T (a subset of variables that are common to both source and target, with an equal probability distribution $P(X)$ in both domains), i.e., a $D_S' \subseteq D_S$ and $D_T' \subseteq D_T$ such that $D_S' = D_T'$, we can define a function $f(\cdot)$ that, given $\mathcal{Y}_S$ associated with D_S, learns on D_S' and predicts $\hat{\mathcal{Y}}_T$ for D_T, given D_T'.

A combination of the to-be-transferred variables $\mathcal{Y}$ and learning function $f(\cdot)$ constitutes a learning task, $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$. In the present application of knowledge transfer, the learning task is invariant for the source and target domains. This means that the same $f(\cdot)$ is applied to the source and target domains: to calibrate function parameters on the former, and to predict $\hat{Y}_T$ on the latter. As with $\mathcal{X}$ and X, we will use Y to denote specific realizations of the variable space $\mathcal{Y}$, i.e. a collection of *n* specific vectors to be predicted in the case of $\hat{Y}_T$, or used to train the learning function in the case of Y_S.

In a given dataset, a specific transfer variable $y \in Y = \{y_1, \dots, y_N\}$ could be either categorical or continuous. Based on this differentiation, the learning function $f(\cdot)$ would respectively involve either a *classification* or *regression* method. Note that according to the naming convention adopted in machine learning, "regression" represents a broad group of methods that go beyond the simple linear or logistic model.

The quality of the prediction of the transferred variables depends on the quality and relevancy of the inputs, X, and the fitness of the learning function $f(\cdot)$. Intuitively it might seem that the more input variables incorporated into $f(\cdot)$, the better the predictions that are generated. However, this intuition collapses in higher dimensions, thanks to the phenomenon commonly known as the *curse of dimensionality* (Bellman, 1961). This concept refers to the exponential inflation of Euclidean hyperspace relative to the unit hypercube as the number of dimensions increases (Keogh and Mueen, 2010). This inflation causes the data to spread out sparsely across the hyperspace and to “drift” towards its edges, all of which leads to a higher variance of the fitted function $f(\cdot)$ and the prevalence of extrapolation over interpolation (Hastie et al., 2009). Possible approaches to abating the curse of dimensionality include variable selection (e.g., stepwise regression) and dimensionality reduction (e.g., principal components analysis).

The fitness of the learning function $f(\cdot)$ can be determined with *cross-validation*, a staple method in the statistical model selection toolbox. Cross-validation resamples the data at random without replacement (unlike *bootstrapping*, which resamples with replacement) to estimate the generalization error of the model $f(\cdot)$ (Du and Swamy, 2013). In a popular variation of *leave-one-out cross-validation*, namely *k-fold cross-validation*, the dataset is partitioned randomly into *k* equally-sized subsets. The learning function $f(\cdot)$ is fitted over the combination of *k-1* subsets (*training data*) while the remaining subset (*test data*) is used to evaluate the performance of the function. The process repeats *k* times on the same partition, with a different subset being used as the test data each time. The prediction errors of $f(\cdot)$ are averaged across the trials to get an unbiased estimate of the generalization error.

Machine learning practice offers a number of different approaches to formulating and estimating the learning function $f(\cdot)$. All primary algorithms used in this study fall into the category of *supervised learning*, namely that $\mathcal{Y}$ exists and is known for the source domain, D_S. In contrast, *unsupervised learning* includes algorithms such as *k-means clustering*, *principal components analysis* (*PCA*), and many others that do not require the prior knowledge of $\mathcal{Y}$ (where, for the examples, $\mathcal{Y}$ is respectively cluster membership and principal component "score") for execution. We implemented a variety of supervised learning algorithms so as to maximize the ability to identify the best ones. Appendix C.1.2 briefly describes the high-level mechanics of the algorithms we used.

4.4 Methodology

4.4.1 Transfer Learning Framework

Our target domain (the recipient of transfer learning) is the NHTS dataset. This domain contains a wide array of disaggregate travel behavior data collected for all 50 states and different land use settings (U.S. Department of Transportation, FHWA, 2009). However, the NHTS sample lacks the attitudinal information that could be instrumental in improving our understanding of travel behavior. It is the purpose of the current study to inform the NHTS dataset with relevant attitudinal data for future use.

Given this objective, a successful source domain (the donor of transfer learning) must include attitudinal variables of interest and should be compatible with the target domain on several levels: First, the two domains should occupy a comparable spatial and temporal continuum to maximize their congruence on *unobserved* attributes. Second, the two domains should possess a pool of *observed* attributes that are equivalent (or can be made equivalent) across domains in their definition, measurement, and marginal distributions $P(X')$. We refer to this pool as *common variables*, denoted X'_S for the source domain and X'_T for the target domain.

With these requirements in mind, we selected the Multitasking Survey of Northern California Commuters (MSNCC) to be the source domain for this study. The MSNCC was administered by the authors between October 2011 and February 2012 (Neufeld and Mokhtarian, 2012). The working cleaned sample contains more than 2,000 observations of commuting adults (this number varies by variable due to scattered, residual item non-response). Attitudes are represented by general opinions (Appendix C, Table C.2), personality traits, multitasking and time use preferences, and transportation mode perceptions. They are measured on 5- and 3-point ordinal scales generally representing degrees of agreement with statements or attributes. In addition to the observed raw data, a series of factor analyses (e.g., Appendix C, Table C.3) was performed to identify the latent constructs underlying each block of interrelated statements (the technical memos describing these factor analyses are available upon request from the authors). Individuals' estimated measurements on these latent constructs are expressed by standardized, continuous Bartlett factor scores. The sign of the factor score indicates individual agreement (+) or disagreement (–) with the latent construct while the magnitude of the score shows the extent of it. Overall, the MSNCC provides a flexible source of categorical and continuous attitudinal data available for transfer learning.

The MSNCC and NHTS data were collected within the same reasonably narrow time window, which makes the two domains temporally comparable. Yet spatially, the domains are not adequately comparable because the geographic area of the MSNCC is a small subset of that of the NHTS. So, unless only a geographically equivalent subset of the NHTS is used (which would dramatically reduce the available sample size and could limit the value of the transferred attitudes for subsequent analysis purposes), extrapolating attribute marginal distributions of the Northern California population (demonstrably not representative of the entire country) to the rest of the target domain could have tenuous validity. However, most attitudinally-rich datasets are geographically limited, and therefore for the purposes of learning more about the circumstances under which these methods are useful, it is pertinent to investigate whether information from a local/regional source can be successfully transferred to a national target. Furthermore, it is possible that although marginal distributions of variables differ between the domains, conditional relationships among multiple variables could be more stable (Babbie, 2010). Accordingly, the analysis reported here used the full nationwide scope of the NHTS dataset (a preliminary analysis showed little impact – specifically, little improvement in the effectiveness of the imputed attitudes – when choosing the California subset as the target for transfer learning).

The source and target domains were identified to have 26 common variables between them (Appendix C, Table C.4). Some of the variables have equivalent meaning and measurement in both domains (for instance, age, gender, race, and household size), while some of the variables require additional manipulation to maximize their congruence (for instance, harmonizing family income categories, determining household life cycle for the MSNCC).

The marginal distributions of the common variables are predominantly different across the two datasets, as is shown in Table C.5 (Appendix C) through visual inspection as well as Kolmogorov-Smirnov and chi-squared tests for continuous and discrete variables, respectively. There are several possible causes for this mismatch: the spatial inequality of domains, varying survey sampling rates, different sampling and data collection strategies, survey non-response, and exclusion of observations due to item non-response. In the transfer learning framework, transductive transfer learning offers specific ways to address the inequivalence of source and target domains, in general, and of the marginal distribution of variables, in particular. For example, assuming that $\mathcal{Y}_T$ is partially known, domain adaptation (Daume and Marcu, 2006) factorizes the marginal distributions of each domain into common and specific parts and uses the three resulting distributions for model estimation and prediction. Alternatively, the iterative proportional fitting procedure could be employed for the key variables to find a set of weights that mitigates the distribution mismatch. However, given the limited time and resources allotted for the present project, the authors decided to leave for future research the process of designing, adding, and evaluating a distribution reconciliation procedure. Nonetheless, readers should keep this caveat in mind while assessing the results presented in this paper.

Additionally to the common variables that are directly available in the MSNCC and NHTS datasets, supplemental land use and "environmental" (i.e., socio-economical aggregates of the immediate surroundings) variables were obtained to aid the transfer learning exercise. As explained further in Section 4.5.1, data from the Decennial census 2010 (U.S. Census Bureau, 2011), American Community Survey (ACS) 2013 (U.S. Census Bureau, 2014), and Smart Location Database 2013 (U.S. Environmental Protection Agency, 2014) were spatially matched to the residential block group of observations in the source and target domains (for the MSNCC, residential locations were reported by the respondents and were therefore available to us, whereas for the NHTS they are made available to researchers upon special request and under strict confidentiality conditions). This augmentation provides supplemental knowledge that could potentially improve the learning function goodness-of-fit. However, such a dramatic increase in the size of the variable space $\mathcal{X}$ prompts a need to deal with the curse of dimensionality effectively to mitigate computational burden and overfitting.

The last piece of the transfer learning framework that has not been defined yet is the learning function $f(\cdot)$, which completes the learning task $\mathcal{T}$ together with the transferred information $\mathcal{Y}$ (attitudes) , $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$. In the context of this study, the learning task is invariant for both source and target domains, i.e., the learning function is estimated on the source domain and applied unmodified to the target domain to predict Y_T. Section C.1.2 (Appendix C) offers a few illustrations of how the specification of $f(\cdot)$ can differ based on inputs, outputs, their interrelationships, form, and preconceived knowledge of all of the above. Each learning function has its strengths and weaknesses with respect to the learning task at hand. The intrinsic uncertainty of which function would perform better in the current setting motivated us to develop a learning-function testing framework as a stage in the project methodology. Applied to the source domain, this framework evaluates the performance of each function by averaging the generalization errors after the 10-fold cross-validation procedure. The learner with the lowest average generalization error (mean-squared error and misclassification error for continuous and categorical dependent variables, respectively) for the test sample is considered the most effective in the current application of transfer learning.

Overall, the methodology of the transfer learning framework developed for this study involves complicated data manipulations, multiple parallel function fittings, and conditional decision-making. It can be succinctly characterized by the following sequence (see Figure 4.1 for a schematic representation).

0. Select and obtain data from the source and target domains: the MSNCC and NHTS (person file).
1. Identify and select common variables across the domains. Reconcile their meaning and units of measurement if necessary.
2. Select and obtain supplemental land use data at the block group level from Census 2010 (summary file, all variables), ACS 2013 (summary file, all variables), and Smart Location Database 2013 (all variables). Expand variable space of the Census and ACS datasets threefold by creating interactions of all variables with the reciprocal of total population and area of a block group, respectively, to create relative, size-independent, land use measures.
3. On each expanded Census and ACS dataset, perform data reduction via principal components analysis (PCA) to extract (unrotated) orthogonal projections of the respective variable spaces.
4. Spatially match the residential locations of the observations in the source and target domains with the Smart Location Database, principal components of the Census data, and principal components of the ACS data, intelligently selecting the number of principal components used from each source. This is important for tuning the computational complexity of the subsequent analyses.
5. On the common variable data matrix X_S' (now including land use data) of the source domain, evaluate the fitness of learning functions by running the 10-fold cross-validation procedure and averaging the generalization errors across the folds.
6. Select tasks (corresponding pairs of an attitudinal dependent variable and learning function) with the lowest average generalization error.
7. For each selected task, estimate the learning function (f_S) on the entire source domain (rather than on the 90% at a time which was used at the cross-validation stage) and use that function to predict the value of the attitudinal variable for

the target domain. Merge the target domain with the transferred knowledge (i.e., predicted attitudes).

Additionally to the description above, during phase #5 we evaluate the performance of the learning functions by investigating both categorical and continuous output variables. At the end of phase #7, we complete the transfer learning process by obtaining the target domain augmented with the knowledge from the source domain.

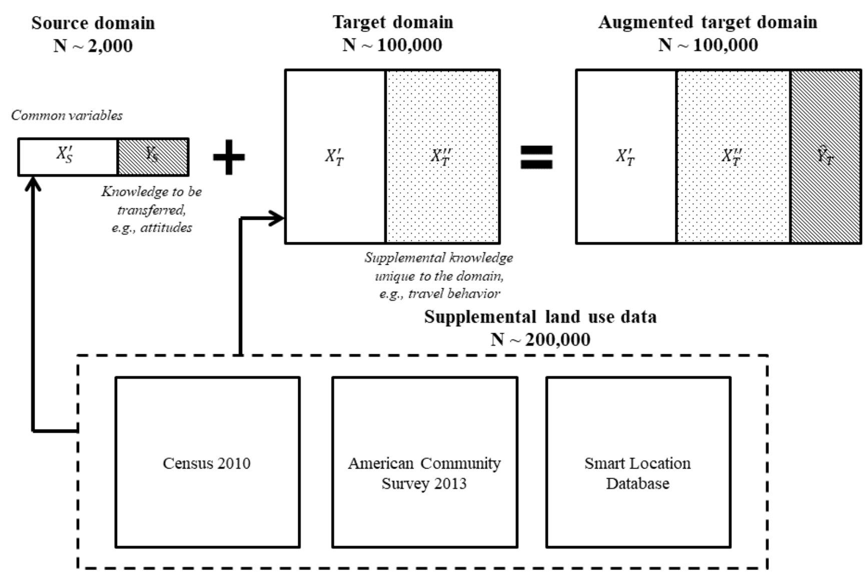

Figure 4.1 – Transfer learning framework (Source: Authors' liberal modification of Fig. 1 of van der Putten et al. (2002))

4.4.2 External Evaluation Framework

Cross-validation is a powerful tool that evaluates how well a prediction function performs if the values of $\mathcal{Y}$ are known. We do not have the benefit of knowing $\mathcal{Y}$ when evaluating the transferred knowledge in the context of the target domain. Nevertheless, it is important to assess the added value of that transferred knowledge. To do that, we propose an external validation framework that is summarized in Table 4.1.

Table 4.1 – External validation framework

Model specification	**Explanatory variable *Y***	**Specification**	**Rationale**
Source Domain			
1	Observed	Best	Benchmark
2	None	Same as *1*, except w/o Y_S	Assess how much explanatory power the observed Y_S has
3	Predicted	Same as *1*	Assess the loss in the goodness of fit of the benchmark model when only the predicted $\hat{Y}_S$ is available
4	Predicted	Best new	Assess how different a model might be from the benchmark, when only the predicted $\hat{Y}_S$ is available and the specification of the model for the true Y_S is unknown
Target Domain			
5	Predicted	Same as *1 & 3*	Assess how well *1* performs within the target domain and with the predicted $\hat{Y}_T$
6	Predicted	Best new	Same as for *4*
7	None	Same as *6*, except w/o $\hat{Y}_T$	See how much explanatory power the estimated $\hat{Y}$ have

In this framework, external validation is realized in the form of a (travel behavior) model implemented on the source and target domain data. To begin, we select a dependent variable that is to be modeled as a function of the rest of the information, including Y (or $\hat{Y}$; note then that in the external validation stages of the analysis, Y and $\hat{Y}$ indicate *explanatory* variables, whereas in the transfer learning stages they were *dependent* variables, or outcomes, of the learning function). Next, we develop models on the source dataset, comparing the outcomes (with respect to quality, fit, and accuracy) across models estimated respectively with Y_S, $\hat{Y}_S$, and neither of those. Finally, we perform a similar analysis on the target dataset, comparing the outcomes obtained across models with and without the transferred (predicted) variables $\hat{Y}_T$.

The first model specification of the framework, which is estimated on the source domain with the observed Y_S, establishes the benchmark of how well the model performs on the observed data. The second model uses the same specification of the previous model except for the exclusion of Y_S from the inputs. Comparing the fits of models *1* and *2* allows evaluating the contribution that Y_S brings to the explanatory power of the benchmark external validation model. The third model has a specification identical to the first one, only instead of the observed Y_S it uses the predicted $\hat{Y}_S$, that is, the output of the learning function $f(\cdot)$ trained on the common variables X'_S of the source domain. The rationale behind this step is to assess how the unavoidably incorrect prediction $\hat{Y}_S$ influences the quality of the validation model. The fourth model seeks for the best new specification, given the *predicted* $\hat{Y}_S$, to assess how the model so obtained might differ from the original model (best specification given the *observed* Y_S). Those differences reflect the data's best compensation for the inaccurate prediction of the transferred knowledge (i.e., some variables may increase or decrease in importance, and other variables may enter the model, to pick up some of the explanatory power lost by replacing the observed Y_S with an imperfect prediction). An assessment of this compensatory mechanism in the source domain can be useful in evaluating the model's performance in the target domain.

These aforementioned four specifications applied to the source domain can provide valuable initial insight into how the external validation (travel behavior) model performs with the observed and predicted data. However, the core of the framework lies in the application of the model to the target domain, in which the observed Y_T is unknown. There, we take the changes in model quality detected in the context of the source domain to be an indication of similar changes in the target domain. Accordingly, the fifth model, which is estimated on the target domain using the same specification as for the first model, has the dual role of establishing a benchmark for the target domain and examining the quality change (compared to that of model *1*) due to the error in the predicted $\hat{Y}_T$. With the search for the best new specification, the sixth model attempts to compensate for the error in the predicted $\hat{Y}_T$ to obtain a better model. Finally, the seventh model specification allows an evaluation of the effects (on the quality of model *6*) of the exclusion of the transferred knowledge.

Although the NHTS is rich in travel behavior variables, the MSNCC is not. However, vehicle ownership is one such variable common to both samples. Accordingly, in this study, we use a vehicle ownership (VO) model for external validation of the transfer learning procedure. We model VO, represented by a count of household vehicles, as a function of variables such as income, number of workers and drivers, and presence of children. Both domains are well-equipped to allow for the specification of a reasonable baseline VO model. Moreover, attitudes have also been found to influence VO (see, e,g., Wu et al., 1999; Cao et al., 2007). Thus, VO is a suitable candidate for the external validation.

4.5 Transfer Learning Results

4.5.1 Data Preparation

The initial MSNCC dataset X_S consists of 1,118 attributes (p) defined for 2,849 observations (n). The person file of the NHTS supplies the dataset X_T of 113 attributes defined for 308,901 observations. Extracting common variables from the datasets shrinks the variable space to 85 attributes for both domains. Since missingness can provide additional knowledge, item non-response on the common categorical variables is coded into an extra dummy variable (=1 if the value of variable x_i is missing, =0 otherwise).

The source domain includes primarily commuters, while the person file of the NHTS dataset contains entire families. To improve the comparability of the domains, preserve commute mode variables for later analyses, and avoid arbitrary predictions for non-commuting populations, we exclude non-commuters from the target domain. After also filtering out observations with item non-response for continuous variables, the target domain shrinks to 112,026 observations.

Before spatial matching, all involved data sources (the source and target domains, ACS, Census, and Smart Location) need to be brought to a common geographic reference. The MSNCC data provides XY coordinates for residential locations. The NHTS spatial

IDs are defined using the 2000 Census block-group boundaries, whereas the three supplemental land use datasets have adopted the block-group boundaries defined for the 2010 Census. We reconcile these two geographies by matching the 2000 Census block-group centroids to the 2010 Census polygons and assigning the corresponding 2010 block-group IDs to the NHTS observations. In this way, all data sources are defined with respect to the 2010 Census geographies.

The original Census and ACS consolidated datasets contain 3,355 and 3,563 variables, respectively. In addition to the absolute numbers, block-group total population and area are used to create two sets of relative measures: share and density – expanding the variable spaces of each consolidated dataset threefold.

Inflating the common variable space of the transfer learning domains by about 20,000 attributes is computationally burdensome and potentially unjustified with respect to prediction accuracy. Moreover, the source domain, which contains just over 2,000 observations, would face the high-dimensionality problem of $p \gg n$, which requires special techniques to treat. For these reasons, we choose to employ a dimensionality reduction method, namely PCA, to decrease the number of attributes while preserving their supplemental knowledge as much as possible. PCA creates successively orthogonal linear combinations (called *principal components*, *PCs*) of the original (intercorrelated) set of variables, in such a way that the first PCs account for the largest shares of the total variance of the original variables. Census and ACS PCs are extracted separately due to the polynomial growth in runtime with the increase in p. In each case, the total number of extracted PCs is $p - 1$: 9,989 and 10,671 PCs for the Census and ACS datasets, respectively. Essentially all variance of the original set of variables is explained by the first

5,084 and 6,187 PCs for the Census and ACS, respectively. These attribute counts are far lower than in the original data, but still unmanageable. As a cutoff, we choose 75% and 50% of the cumulative variance explained, corresponding to 120 and 76 PCs, for the Census and ACS, respectively.

The Smart Location dataset contains 117 variables, which cover such attributes as demographics, employment, density, diversity, design, transit, and destination accessibility (the full data dictionary is available in Ramsey and Bell (2014)). The relatively small variable space of this supplemental land use dataset allows spatially matching the data without requiring a dimensionality reduction step. After the supplemental land use datasets are spatially matched based on the residential location, the dimensions of the domains are 2,352×379 and 91,666×380 for the source and target, respectively.

The final step of data preparation is to augment the common continuous variables in both domains by replacing them with their *natural cubic splines* (degrees of freedom = 3). This process is called *basis expansion*. Using splines is a relatively simple way to allow for non-linearity in relationships in additive models. However, a downside of expanding the basis is the inflation of the continuous variable space by the factor of the degrees of freedom. After replacing continuous explanatory variables (including PCs) in the transfer learning domains with their cubic splines, the common variable space of source and target datasets expanded to 968 attributes.

4.5.2 Best Learning Function Search and Selection

For the source domain (2,352×969), the search for the best-performing learning function is accomplished by measuring the generalization error and averaging it over a 10-

fold cross-validation (CV) routine. We explored two different approaches to predicting attitudes. In the first approach we focused on directly predicting the continuous-valued factor scores that had been previously computed for the source domain (see Table C.3 in Appendix C for examples of factor content). In the second approach, we first predicted the ordinal responses to individual attitudinal statements (such as those in Table C.2 in Appendix C), and then factor-analyzed those predicted responses.

The search for best learning function is performed separately for the continuous (attitudinal factor scores) and categorical (attitudinal statements) dependent variables, respectively instances of the regression and classification problems described in Section 4.3.2[17]. Recall that in the present discussion, the "dependent variable" refers to the attitudinal variable being predicted ($\hat{y}_i$), in contrast to the dependent variable (in our case, vehicle ownership) of the model introduced for external validation in Section 4.4.2, in which the observed (y_i) and predicted ($\hat{y}_i$) attitudes are *explanatory* variables. This subsection describes phase #6 from the transfer learning methodological sequence defined in Section 4.4.1.

4.5.2.1 Regression Problem

For the regression problem, Table 4.2 presents selected generalization errors (the mean squared errors, MSEs) obtained for the continuous dependent variables given the corresponding learning functions ("learners"), i.e., regression tasks (the full results can be found in the Appendix C, Table C.8). We tested eleven different learners: random hot deck

[17] Appropriate algorithms, which are capable of handling either or both types of problems (regression and classification), are described in Section C.1.2 (Appendix C).

(RHD), assigning the mean value, forward stepwise linear regression, classification and regression tree (CART), evolutionary regression tree, recursive tree, bagging, random forest, LASSO regression, support vector machine (SVM), and AdaBoost. Among these, LASSO regression (linear regression kernel) shows the best performance by having the minimum generalization error for all $\mathcal{Y}$ variables, except for the *Time-pressure – reality* factor score. On average, the LASSO MSE is 0.894 (which, taking the square root, represents about one standard deviation off the observed value) across the nine dependent variables shown, which is an 11% and 55% improvement over assigning the mean value and RHD, respectively. The RHD learner, as expected, demonstrates the worst performance with an MSE of 1.986 (1.4 standard deviations off the observed value). Mean value assignment, the other learner that is free of conditional assumptions, outperforms only two methods: RHD and forward stepwise linear regression. The latter performed relatively poorly because of the increased prediction variance due to overfitting at the training stage. This is especially interesting since LASSO is also a linear regression method with a variable selection routine. The difference between the two is that LASSO has a built-in cross-validation procedure (done for each fold of the higher-order CV) to prevent overfitting.

Table 4.2 – Selected cross-validation results for the regression problem

Variable	**Best learner**	**Lowest MSE**	**Mean assignment MSE**	**ΔMSE (mean assignment vs. best learner)**
Pro-transit	LASSO regression	0.757	0.993	–0.236
Travel is wasted time	LASSO regression	0.985	1.001	–0.016
Pro-technology	LASSO regression	0.951	1.017	–0.066
Commute benefit	LASSO regression	0.898	1.008	–0.110
Time pressure – reality	Evolutionary regression tree	0.994	1.009	–0.015
Time pressure – preference	LASSO regression	0.936	0.994	–0.058
Pro-active transportation	LASSO regression	0.789	1.009	–0.220
Satisfaction	LASSO regression	0.976	1.004	–0.028
Pro-density	LASSO regression	0.748	1.005	–0.257

The prediction performance of the tasks varies across the dependent variables. *Pro-density*, *pro-transit*, and *pro-active transportation* factor scores are predicted by LASSO regression with an MSE below 0.8. For these variables, the greatest deviation (improvement) from the mean value assignment is achieved: ΔMSE is above 0.20. *Commute benefit* has a slightly worse prediction success with a generalization error of 0.898 (ΔMSE=0.11). The other five variables show substantially less improvement over the mean value assignment method, with ΔMSEs below 0.07. It stands to reason that the observed distribution of the generalization error is affected by the knowledge content (i.e., relevance) of the common variables used for prediction. The heavy prevalence of land use inputs in the source domain caused the learning functions to explain relatively well the attitudes associated with built environment attributes. Specifically, commuters who score high on *pro-density*, *pro-transit*, *pro-active transportation*, and *commute benefit* attitudes are more likely to live in denser neighborhoods with more transit, bicycling, and walking options due to residential self-selection, a phenomenon that prominently features in recent literature (Cao et al., 2009).

4.5.2.2 Classification Problem

In a regression problem, trying to predict a continuous variable could produce an unsatisfactorily large generalization error if variables that strongly influence the error's bias and variance components are unobserved and unaccounted for. In this situation, solving a classification problem, where the goal is to predict to which one of a (usually) small number of predefined categories to which the observation belongs, could mitigate the role of the unobserved inputs and decrease the influence of the error's components. Additionally, classification problems require certain changes in the algorithm of the learning functions, or the use of completely new learners, which, potentially, might better capture the associations existing in the data. Finally, in using predicted attitudinal items as inputs to a factor analysis, we speculate that random errors associated with predicting each single item could partially counteract each other and result in predicted factors that are more accurate (closer to the "observed" factor scores previously computed from the observed attitudinal items) than those predicted directly as just described. Accordingly, we also performed the prediction of individual items with ordered categorical responses. However, the cross-validation accuracy results obtained in this way (see Section C.2, Appendix C) were apparently not superior to those obtained for the regression problem.

4.5.2.3 Comparison of the Outcomes of the Regression and Classification Problems

To compare the results of continuous and categorical dependent variable prediction, we investigate how the direct prediction of factor scores (regression problem) fares relative to the prediction of the raw statements (classification problem) with subsequent factor analyses of the predicted data. In all factor analyses we use the original method: principal axis factoring with oblimin rotation.

While more details are available in Malokin et al. (2017b), here, we summarize them as follows. Comparison of (1) the factor scores obtained from multiplying the common factor score coefficient matrix by the various sets of predictions to (2) the scores originally computed using the observed attitudes shows consistently high correlations for the same constructs identified in the regression problem: *pro-density*, *pro-transit*, and *pro-active transportation*. However, for the most part the highest correlations obtained in this step are still worse (lower) than those obtained from the results of the regression problem. We conclude that at least in this instance, the direct prediction of factor scores is better (and more straightforward) than the two-stage process of predicting individual statements and then factor-analyzing them. Nevertheless, it is still potentially useful to have access to the predicted attitudinal statements, for situations where individual items may be of specific interest, and/or do not load heavily on any factor.

4.5.3 Transfer Learning

For the transfer learning procedure, we apply the learning task to the entire source domain (as opposed to the CV procedure, which uses only a subset of the domain), corresponding to phase #7 of the methodological sequence defined in Section 4.4.1. The common variable space contains all variables described in Table C.4 and the land use data described in Section 4.5.1. The learning task consists of LASSO regression as the learning function and attitudinal dependent variables (sequentially paired with the learner).

However, even though the regression problem is shown to be better suited in the setting of the current study (Section 4.5.2), it is not computationally-burdensome to carry out the classification problem also. Accordingly, using LASSO regression with, respectively, linear regression and MNL kernels for the regression and classification tasks, we estimate the learner for each transferred variable on the source domain and apply this learner to the target domain. At the end of the transfer learning procedure, the target domain variable space receives 9 continuous and 39 categorical attitudinal variables defined for 91,362 observations (respondents in the NHTS person file). Table 4.3 presents selected distribution parameters of the transferred continuous variables, which we briefly discuss here (Table C.6 and Table C.7 of the Appendix C contain similar information on the observed and predicted attitudes for the source domain.).

Table 4.3 – Descriptive statistics of the transferred continuous attitudes for the NHTS dataset (N=91,362)

Variable	**Number missing**[a]	**Mean**	**SD**	**Median**	**Min**	**Max**	**Skew**	**Kurtosis**
Pro-transit	1	–0.31	0.29	–0.34	–3.21	3.26	1.41	5.33
Travel is wasted time	0	0.01	0.10	0.01	–0.58	3.84	5.03	124.49
Pro-technology	5	0.00	0.25	0.01	–4.02	4.97	0.94	15.45
Commute benefit	0	0.10	0.34	0.10	–3.49	3.04	–0.66	7.96
Time pressure – reality	0	–0.05	0.11	–0.05	–0.77	0.76	0.01	0.02
Time pressure – preference	0	–0.05	0.19	–0.04	–1.86	0.99	–0.21	0.13
Pro-active transportation	5	–0.39	0.29	–0.42	–1.94	4.98	1.77	16.57
Satisfaction	1	0.12	0.21	0.14	–1.65	4.17	–0.08	12.85
Pro-density	3	–0.42	0.45	–0.46	–1.53	3.28	0.70	1.17

[a] Predicted values beyond +/–5.0 are coded as missing. Since the learning function is trained on a smaller sample, prediction for some observations in the NHTS sample (which is larger, more heterogeneous, and with a greater chance of extreme input values) could be a result of extrapolation rather than interpolation. The former is known to be more unstable and to produce unrealistic outcomes.

With respect to the continuous attitudes (factor scores), we first note that attitudes per se do not have an "absolute" zero point – they can only be measured relative to some arbitrary benchmark. Accordingly, in the source domain, the attitudinal factor scores were standardized variables, so that each of their means were zero, and standard deviations equal to one (for the MSNCC dataset, N=2,849). This effectively makes the Northern California sample of the source domain the benchmark against which the national sample of the target domain is measured. A mean factor score that is close to zero in the target domain signifies that on average, the national sample holds an attitude similar to that of Northern California. With that in mind, we can see from Table 4.3 that based on the nationwide predicted factor scores, respondents are considerably less pro-transit, pro-active transportation, and pro-density than those in the Northern California sample are (while national respondents are comparatively somewhat more satisfied with life and job, and view the benefits of commuting somewhat more positively). Although this may not be surprising in terms of Northern California stereotypes, it is important to keep in mind that the MSNCC sample is deliberately enriched with non-drive-alone commuters (Neufeld and Mokhtarian, 2012), and as such, in raw form it is not even representative of Northern California.

It is also important to note that all the standard deviations of the predicted scores are markedly smaller than one. While in *theory* this could indicate that attitudes in the Northern California sample are considerably more variable in the aggregate than are attitudes nationwide[18] (which could be another consequence of the choice-based sampling

[18] Of course, the *range* of attitudes in Northern California will be encompassed by the range for the nation that contains Northern California, but (loosely speaking) if in the national sample extreme attitudes are a *smaller share* of the total, the standard deviation will be smaller. On the other hand, it can be argued that choosing a source sample to have greater variability than the target sample (i.e., choosing it to overrepresent more extreme opinions) is not necessarily a bad thing: a more variable source can draw on more knowledge

strategy), it is presumably to a much greater extent a reflection of prediction error: given that most sources of variability in attitudes are unmeasured, the learning function will tend to make predictions that do not vary far from the sample mean.

This supposition is strongly supported by a comparison of Table C.6 (descriptive statistics for the observed scores in the source sample) and Table C.7(descriptive statistics for the predicted scores – also in the source sample): whereas standard deviations (s.d.s) of the observed scores are all close to one (by design), standard deviations of the predicted scores are never higher than 0.48. Not surprisingly, the three predicted attitudes with the largest standard deviations (where, in this case, a s.d. that is larger – therefore closer to that of the observed attitude – is better, suggesting that the learning function is better at explaining the natural variability of the factor) are pro-density (0.48), pro-transit (0.46), and pro-active transportation (0.45) – the three best-predicted attitudes in this analysis (Section 4.5.2.1). Commute benefit (0.34) comes in fourth, also in keeping with its predictability.

Comparing the standard deviations of the source domain's predicted factor scores (Table C.7) to those of the target domain's predicted scores (Table 4.3) offers further insight: for most of the nine factors, the s.d.s are nearly equal, whereas for two of the better-predicted factors (pro-transit and pro-active transportation), they shrink by about a third in the target domain, indicating that the cross-sample transferred factor scores are substantially less variable than the own-sample predicted ones are. Interestingly, the *ranges* of observed and predicted factor scores display a different pattern: the ranges vary between

---

in predicting a less variable target, than if a more homogeneous source were attempting to predict for a more variable target.

5 and 7 for the observed continuous attitudes (Table C.6); they shrink at least twofold (up to ninefold in some cases) for the predicted attitudes in the source dataset (Table C.7); and they take on more variable amplitudes – larger as well as smaller (2-9) – for the predicted attitudes in the target domain (Table 4.3). While the larger ranges in the latter instance suggest a promising departure from the mainly homogeneous predictions seen in the source dataset, it might be an artificial effect created by the learning function struggling with extrapolation in the context of the greater sample heterogeneity of the NHTS.

In sum, these statistics offer a useful reminder of the relativity of attitudinal measures. It is clear that the source sample differs substantially from the target sample in its distribution of the target variables. As discussed in Section 4.4.1, in future work the source sample (if not initially drawn from the same population as the target sample, which would be preferable) can be weighted to be more representative of the target in terms of the common variables, which should reduce or eliminate these differences. In the meantime, future users of the scores predicted for the national target domain may wish to re-standardize them. This would at least establish the national mean as the benchmark, although it would not resolve the lower variability in predicted values.

4.6 External Validation Model Results

For VO models, the dependent variable, number of household vehicles (*HHVEHCNT* in the NHTS data dictionary), is defined in both domains as a count of motorized vehicles that a household owns. For the pool of potential explanatory variables, we select attributes common to both domains that have been used extensively in the literature and proven to influence VO. (Note that the same variables, albeit a superset of them, are used in the transfer learning exercise.) This pool includes race, gender, age, education, immigrant status, full/part-time work status, occupation, conditions preventing driving/taking public transit, household income, presence of children, number of children, number of drivers, number of workers, interaction between number of workers and number of drivers, distance to work, and land use variables. For greater interpretability, selected land use variables (population, employment, and network densities) are sourced from the Smart Location Database, instead of using the mechanically derived and conceptually abstract Census and ACS principal components described in Section 4.4.1.

In addition to this list, the pool of explanatory variables includes attitudes, represented by the three latent constructs that showed the lowest generalization error during the cross-validation step: *pro-transit*, *pro-active transportation*, and *pro-density*. We believe that these attitudes should capture effects associated with transportation mode preference and (through residential self-selection) availability, thus influencing the household's VO.

There are several conventional ways a VO model could be specified, including using linear regression, Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial, ordinal response, or multinomial discrete choice (including nested) functional forms. For this study, we choose linear regression due to the interpretability of its standard goodness-of-fit measure. Furthermore, our testing found that other formulations produced very similar results.

Table 4.4 shows the resulting goodness-of-fit measures, together with coefficient signs and significance levels, for the seven linear regression model specifications that constitute the external validation framework. Model *1*, a benchmark, is estimated on the source domain with observed attitudes. All coefficients have the expected sign; in particular the attitude coefficients are strongly significant and negative, indicating that the more pro-transit, pro-active transportation, and/or pro-density respondents are, the fewer vehicles their households will tend to own. The adjusted R^2 of this specification is 0.45 – an indication of a reasonably well-specified model. Model *2*, obtained by setting the attitudinal coefficients to zero, has an adjusted R^2 of 0.42, which signifies a 0.03 (8.0%) "model lift" (improvement in fit) attained by accounting for the three attitudinal constructs in Model *1*. Using the same benchmark specification but replacing observed with predicted attitudes (Model *3*) fits the data even slightly better ($\Delta R^2 = 0.0015$). In this specification, even with the *Pro-active transportation* coefficient being insignificant (yet still negative), the three predicted attitudes combined are able to explain the variance of the dependent variable better than the originally "observed" attitudes. One possible explanation for this could be the knowledge from the tens of thousands of variables used to predict the attitudes (see Section 4.4.1). I.e., this multitude of "hidden" variables, which are not present in Model *1*, evidently contains a small amount of explanatory power above and beyond the variables that do appear in that model. Model *4* reinforces this empirical result by showing that an optimized (newly-specified "best") model with predicted attitudes improves over the previous two ($R^2 = 0.46$).

Table 4.4 – External validation framework results: linear regression VO model results

Model specification[a]	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Domain	Source	Source	Source	Source	Target	Target	Target
Attitudes	Observed	N/A	Predicted	Predicted	Predicted	Predicted	N/A
Specification	Best	1 w/o atts.	1	New best	1	New best	6 w/o atts.
Adjusted R-squared	**0.4544**	**0.4209**	**0.4559**	**0.4565**	**0.3849**	**0.3894**	**0.3848**
Variable[b]							
Intercept	+++	+++	+++	+++	+++	+++	+++
Pro-transit	---	0	---	---			0
Pro-active transportation	--	0			---	---	0
Pro-density	---	0	---	---	---	---	0
HH_HISP	0	0	0	0	0	---	--
HH_RACE: Black	--	--	--	---	---	---	---
HH_RACE: Asian	0	0	0	-	0	--	--
HH_RACE: Multi	--	--	---	---		0	0
HH_RACE: Other	-	-	-	--		0	0
HHFAMINC: $0-25k	---	---	---	---	---	---	---
HHFAMINC: $25-50k	---	---	---	---	---	---	---
HHFAMINC: $50-75k	---	---	---	--	---	---	---
HHFAMINC: $75-100k	---	--	---	0	---	---	---
HHFAMINC: >$100k	0	0	0	+++	0	0	0
Was born in the U.S.?	++	++	++	+	+++	+++	+++
Condition preventing using public transit	0	0	0	0	0	---	---
EDUC: less than HS degree	0	0	0	0	0	+++	+++
EDUC: HS degree	0	0	0	0	0	+++	+++
EDUC: less than BS/BA degree	0	0	0	0	0	+++	+++
OCCAT: service	0	0	0	0	0	+++	+++
OCCAT: clerical	0	0	0	0	0	+	
OCCAT: manufacture	+++	+++	++	++	+++	+++	+++
OCCAT: professional	0	0	0	0	0	+++	++

Table 4.4 (continued)

Model specification[a]	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Domain	Source	Source	Source	Source	Target	Target	Target
Attitudes	Observed	N/A	Predicted	Predicted	Predicted	Predicted	N/A
Specification	Best	1 w/o atts.	1	New best	1	New best	6 w/o atts.
Adjusted R-squared	**0.4544**	**0.4209**	**0.4559**	**0.4565**	**0.3849**	**0.3894**	**0.3848**
Variable[b]							
R_SEX	**0**	**0**	**0**	**0**	**0**	--	
SELF_EMP	**0**	**0**	**0**	**0**	**0**	+++	+++
Works full time?	**0**	**0**	**0**	**0**	**0**	---	---
DRVRCNT	+++	+++	+++	+++	+++	+++	+++
WRKCOUNT	+++	+++	+++	+++	+++	+++	+++
R_AGE	**0**	**0**	**0**	**0**	**0**	-	
DISTTOWK	+	+	+	+	+++	+++	+++
Population density	-	---		**0**	---	---	---
Activity density	**0**	**0**	**0**	**0**	**0**	---	---
Jobs per HH	-	--	-	**0**		**0**	**0**
Road network density	**0**	**0**	**0**	**0**	**0**	---	---
Jobs within 45 mins	---	---		--	---	**0**	**0**
Number of children	-		--	--	---	---	---
Presence of children	**0**	**0**	**0**	**0**	**0**	---	--
*DRVRCNT*WRKCOUNT interaction*	---	---	---	---		**0**	**0**

[a] Numbering corresponds to Table 4.1.
[b] Model coefficients are represented by their sign (+ for positive, – for negative, blank for insignificant) and significance (one sign for p < 5%, two signs for p < 1%, three signs for p < 0.1%). Zeros indicate the coefficient's absence from the model specification.

The benchmark model specification applied in the target domain context (Model *5*) shows a loss of significance of the *pro-transit* coefficient and changes its sign to a counterintuitive positive one. However, given the national coverage of the NHTS, it is not surprising that a pro-transit attitude plays a lesser role outside the relatively small number of transit-oriented areas. This explanation is further corroborated when the subset of urbanized regions with well-developed transit is isolated from the target domain for model estimation purposes. For example, Model *5* estimated only on the State of New York (results not shown) produces a highly significant and negative *pro-transit* coefficient. Returning to the model estimated on the full target domain, compared to the first specification, the goodness-of-fit measure is lower ($R^2 = 0.38$), which could be another effect of the greater heterogeneity in the nationwide sample. Model *6* is a product of the search for the best specification in the context of the target domain. It slightly improves over Model *5* ($\Delta R^2 = 0.0045$), with the attitudinal effects demonstrating the same pattern (i.e., the *pro-transit* coefficient is not statistically significant, and is positive). Finally, Model *7* (when compared to Model *6*) answers the main question of the value of the transferred knowledge (attitudes) for future analyses. The exclusion of attitudes from the model specification results in a drop in the goodness-of-fit measure of 0.0046, or conversely, adding the three attitudinal latent constructs to the VO model specification increases the variance explained by 1.2%.

At first glance, the model lift of 1.2% is rather weak. However, it is useful to consider what variables have been used for the knowledge transfer and external validation processes. By design, the inputs of both the LASSO regression learning function and the VO model are drawn from the partly overlapping subsets of the common variables. With the same socio-economic, travel behavior, and selected land use variables being used in both of these linear-in-parameters functions, the predicted attitudes have little remaining explanatory power to offer beyond that of the other variables in the VO model. When this circularity is removed, i.e., when, for example, only land use variables are used in the transfer learning step and only socio-economic and travel behavior variables (together with the predicted attitudes) are used in the external validation step, the model lift rises to 5.4%, or $\Delta R^2 = 0.0197$, much closer to the difference between Models *1* and *2*. Although this obtains a more reassuring performance for the transferred attitudes, it is achieved at the cost of omitting land use explanatory variables – known to be relevant to predicting VO – from the VO model.

Table 4.5 paints a more comprehensive picture of the competition for explanatory power, as VO model goodness-of-fit measures are cross-tabulated with respect to the groups of variables used in the transfer learning and external validation model specifications. Focusing first on the rows, the table shows that when blocks of variables are entered singly, the socio-economic block delivers the most sizable jump in R^2 (0.36) for the VO model, while attitudes by themselves are quite modest in predicting household vehicle ownership (R^2 ~ 0.04-0.07). When separately combined with socio-economic variables, the attitude and land use variable blocks each enhance the goodness-of-fit measure by approximately 0.02. An even slighter further increase is demonstrated when all three groups of variables are used together to model VO, indicating the diminishing returns of including correlated explanatory variables.

Turning to the columns, it can be seen that which blocks of variables are used to predict attitudes also influences the goodness of fit of the VO models. When entering the blocks singly, using only land use variables to predict attitudes yields better-fitting VO models than using only socio-economic variables – which, again, is not surprising in view of the land-use-related nature of the attitudes in question. Using both socio-economic and land use variables as predictors for the attitudes further improves the VO models, but only very little beyond what having the land use variables alone delivers. Overall, as data availability increases for both the transfer learning and external validation models, the latter benefits by having a higher goodness-of-fit measure, but the incremental benefits are modest.

Table 4.5 – Goodness-of-fit (R^2) of VO models in the target domain (NHTS) by VO model specification and LASSO regression learning function inputs

Vehicle ownership is a function of …	Attitudes are a function of … Socio-economic variables only	Land use variables only	Socio-economic & land use variables
Attitudes only	0.0466	0.0408	0.0659
Socio-economic vars. only		0.3572	
Attitudes & socio-economic variables	0.3655	0.3796	0.3797
Land use & socio-economic variables		0.3764	
Attitudes & land use & socio-economic variables	0.3808	0.3844	0.3851

4.7 Summary and Conclusions

In this paper, we have developed a transfer learning-based framework for enriching one domain (consisting of a data matrix and the probability distribution of the variables) with knowledge obtained from other related domains. At the heart of this framework lies the process of identifying the set of variables common across the datasets and training a learning function that performs the knowledge transfer from the source domain (in which the transferred variables of interest are *observed*) to the target domain (where the transferred variables of interest are *statistically inferred*). To evaluate the performance of the transferred knowledge, we have also proposed an external validation framework. This framework employs a model, external to the transfer learning process, which is estimated using the transferred knowledge as inputs. Thus, the external validation model provides empirical insight into how valuable the transferred knowledge is to the target domain.

The transfer learning framework of this paper is broadly applicable to many types of knowledge. The specific aim in this study was to use the framework to enrich the National Household Travel Survey data with attitudes transferred from another dataset. In our application, the *pro-transit*, *pro-active transportation*, and *pro-density* attitudinal factor scores showed the lowest generalization error (using the LASSO learner) and the greatest improvement over the benchmark (assignment of the mean value). The external validation framework was implemented by using a vehicle ownership linear regression model estimated on the source and target domains with observed and predicted attitudinal factor scores. The external validation revealed that in the source domain the observed attitudes account for an 8.0% model lift, and in the target domain the predicted attitudes account for a 1.2% model lift.

The latter modest result can be explained by the widely overlapping variable space that was used in both the transfer learning and external validation frameworks, which forced the predicted attitudes to compete with their predictors for explanatory power within the same external validation model. This effect was aggravated by the linear-in-parameters nature of the functions used in both frameworks, which created a more straightforward substitution and "double counting" patterns among the same variables. If the dependency on the same variable space for both frameworks is broken (i.e., in this instance, when only land use variables are used in the transfer learning step and only socio-economic and travel behavior variables, together with the transferred attitudes, are used in the external validation step), the target domain shows a model lift of 5.4% when the attitudinal factor scores are included. Excluding land use variables as direct predictors of VO, however, has problems of its own, as discussed in Section 4.5.

The benefit and cost of strict separation between inputs of the two frameworks is, perhaps, the most important finding of this study. Arguably, the transfer learning process could be viewed as a dimensionality reduction exercise that integrates a vast input variable space into a handful of attributes that gain their definition and meaning from the original knowledge (dependent variables) to be transferred. In light of these findings, we recommend applying the transfer learning framework to supplemental datasets (e.g., land use, marketing, socio-economic environment data, etc.) that offer reasonable ways to match them to the source and target domains, have strong associations with the transferred knowledge, and serve as valuable informational supplements to future analyses.

This study is far from conclusive. We highlight six important limitations and convenience/necessity shortcuts that warrant further investigation:

1. **Domain adaptation**. Achieving spatial and temporal equivalence between the source and target domains could be a difficult task, given the heterogeneity that exists in data acquisition. Thus, more effort should be dedicated to researching methods of assuring comparability among domains, including reconciling the marginal distributions of the common variables.
2. **Knowledge transfer functions**. The machine-learning field continues developing more advanced and sophisticated methods for more accurate and reliable predictions.

3. **Obtaining a variety of data**. Some potentially fruitful sources of additional common variables include marketing data, credit card transactions information, economic and business aggregates, social media activity, geolocation data, and so on.
4. **Evaluating performance of tasks given the available data**. The three main components of the transfer learning framework are the input (common) variables, the learning functions, and the output (transferred) variables. Options for each of these offer a large number of possible combinations. A more systematic investigation / mapping of generalization errors for various combinations of inputs, dependent variables, and learners is needed.
5. **Evaluating external validation framework**. Similar to the previous point, numerous combinations of components that come into play for the external validation framework need to be further investigated. Effects of knowledge recycling (or its absence) on model lift and different kinds of external validation models are pertinent topics for future research.

In a world where more than 2.3 million terabytes of data are generated every day (VCloudNews, 2015) – and this rate is growing rapidly – the problem of distilling data into humanly-tractable and actionable knowledge is paramount. With the current transfer learning methodology, we arrived at a dimensionality reduction technique of predicting transferred variables as a surrogate for the common variable space. We see this approach as an effective way of treating the $p \gg n$ problem with an advantage of substituting vast variable spaces with meaningful transferred variables, which are suitable for subsequent classical statistical analyses and decision-making processes. Nevertheless, there is much left to learn and improve.

4.8 Acknowledgements

This study was primarily funded by a grant from the Transportation Energy Evolution Modeling division of the Oak Ridge National Laboratory. Additional funding came from the Center for Teaching Old Models New Tricks (TOMNET), a University Transportation Center sponsored by the U.S. Department of Transportation through Grant No. 69A3551747116. The authors would like to thank the Travel Monitoring Division (HPPI-30) of the Federal Highway Administration, and personally Jasmy Methipara, Adella Santos, and Tim Reuscher, for providing access to the confidential part of the National Household Travel Survey.

CHAPTER 5. CONCLUSIONS

5.1 Summary of the Research

In this work we accomplished the main research objective of contributing to the improvement of regional travel behavior models by investigating the influence of understudied behavioral drivers and by increasing the availability of attitude-based insights to those models. To achieve this goal we proposed the *Production-Transduction* framework, which, generally, comprises two steps. First, at a practical scale, new data that can be used to establish relationships between previously understudied (in a specific context) variables and travel behavior is collected and analyzed (i.e., *production*), and, second, separate datasets are bridged to spread the "exclusive" data more widely (i.e., *transduction*).

For the empirical application of the *Production-Transduction* framework we investigated the effects of travel-based multitasking on mode choice and the value of travel time (*Production,* Chapters 2 and 3), and developed an approach for transferring attitudinal data from a small regional dataset to a large national sample (*Transduction*, Chapter 4). The two steps are connected by the usage of an attitudinally-rich travel behavior dataset of Northern California commuters. Although in the present application the targeted variables diverged between the two steps, that need not be the case in future applications. The two steps of the framework can be executed separately or together, and the framework per se offers a fully general methodology with innumerable potential applications.

The empirical results of the present application are also meaningful in their own right. In particular, in Chapter 2, the multinomial logit estimation results showed that multitasking is significant to mode choice in three ways as *multitasking mode perception* (for all alternatives), *general preference towards multitasking* (*polychronicity*, for the shared ride alternative), and *propensity to use a laptop/ tablet/ notebook* (for all alternatives). All model coefficients associated with these explanatory variables are positive, indicating that the studied multitasking factors increase respective utilities of the alternatives and make the probability of the mode to be chosen greater. The scenario analysis demonstrated that the contribution of multitasking to the currently observed aggregate mode shares in the studied region is modest yet non-negligible, as driving alone commute shares would be 1.5 percentage points higher if usage of a laptop/ tablet/ netbook were unavailable. Interestingly, being more multitasking-friendly, autonomous vehicles could bring a virtually similar increase in the current shares of the single occupancy alternative if their passengers have the opportunity to use a laptop/ tablet/ notebook during the commute. Therefore, the diffusion of autonomous vehicle technology might pose a threat to collective modes such as public transit and contribute to increases in VMT and urban sprawl.

In Chapter 3, we analyzed the commuting travel behavior of millennials vis-à-vis older adults with respect to the impact of activities while traveling on mode choice, value of travel time savings (VOTTS) and willingness to pay for productive travel multitasking. Our findings indicated important implications for planning and modeling applications. Specifically, we could not find any significant influence of socio-economic variables on the commute mode choice of young adults. Millennials' heterogeneity of travel behavior was explained by mode attributes, attitudes (mode-specific perceptions, general travel-related opinions), and travel multitasking propensity. A sensitivity analysis indicated that the failure to account for activities conducted while traveling strongly biases the estimated travel time coefficients for millennials, overstating their VOTTS. Our results show that younger adults value information and communication technologies (ICT) usage more highly because it gives them the ability to make travel more tolerable, and they are willing to pay more for it. This result potentially amplifies the findings from Chapter 2, as millennials are assuming an increasingly influential role in shaping aggregate travel behavior patterns (on account of being the most populous cohort).

In Chapter 4, we have developed a transfer learning framework for enriching one dataset with variables obtained from another related dataset. To evaluate the performance of the transferred variables, we have also proposed an external validation framework. This framework employs a travel behavior model, external to the transfer learning process, which is estimated using the transferred variables as inputs (among other relevant explanatory variables). Thus, the external validation model provides empirical insight into how valuable the transferred variables are to the enriched dataset. By using an attitudinally-rich dataset of Northern California commuters (analyzed in Chapters 2 and 3) as the source of general transportation-related attitudes and the NHTS 2009 as an enrichment target, we applied machine learning methods and techniques to predict the transferred variables in the context of the nation wide dataset. The external validation revealed that the predicted attitudes account for a 1.2-5.4% model lift. Overall, the proposed transfer learning framework offers an inexpensive and fast way to supplement existing datasets with novel variables that are relevant for specific purposes, to benefit travel demand forecasting, transportation planning, and decision making as transportation systems face unprecedented changes and challenges.

5.2 Research Limitations

Among the major research limitations, we recognize that this research is noticeably dependent on the Multitasking Survey of Northern California Commuters, data for which was collected within a relatively small geographic region in 2011 and 2012. As such, the transferability between contexts of the spatial, temporal, and content-wise characteristics of the dataset are not ideal. However, the proposed methodologies for enriching our understanding of travel behavior with attitudinal variables are robust and capable of being implemented in various contexts.

This research investigates only a tiny sliver of the panoply of travel-based multitasking, i.e., commuters who used a laptop/ netbook/ tablet on their primary commute mode. The number of possible travel-based activities, travel modes (and ways to chain them), and travel purposes offer a much more diverse landscape that is of interest for future analyses.

Additionally, the proposed transfer learning framework (Chapter 4) allows for potential endogeneity, as the explanatory variables that are used for transferring variables between datasets could be used for the subsequent modeling jointly with the transferred variables. This issue could be resolved either by restricting the sets of input variables for each modeling step (i.e., transferring and modeling) to be non-overlapping or using non-leaner transfer functions (and adding stochastic noise to the prediction).

The machine learning methods that were implemented in the transfer learning exercise were used in a fashion that pursued the breadth rather than depth of their application due to the exploratory nature of this work. Finetuning of model hyperparameters would, most likely, improve the performance metrics, as would the acquisition of a wider set of common variables.

5.3 Directions for Future Research

With respect to travel multitasking, it would be of interest for future research to investigate the relationships of other multitasking propensity factors to mode choice, likely using structural equation models to test specifications that allow for multiple directions of causality, e.g., allowing activities conducted while traveling to be influenced by mode choice and influence mode choice simultaneously. It is also of interest to explore population heterogeneity (e.g., Shaw et al., 2018 looked at the reported benefits and disadvantages associated with the activities conducted while commuting), and undertake international comparisons of the impact of multitasking on the perception of travel utility by mode, and mode choice (through additional data collection in other regions). Additionally, as ICT technology and its percolation changes with time, conducting a follow-up study to analyze the difference between effects now and then could be of benefit.

As the initial study suggested, the transfer learning exercise benefits greatly from input data availability. Identifying and acquiring the supplemental data to expand the pool of the common variables used in transfer learning is potentially the most fruitful avenue of further research. Accordingly, our research group has obtained a rich targeted marketing dataset to investigate how machine learning methods can leverage data that may appear to be irrelevant to transportation (but which may in fact provide substantial information about attitudes and lifestyles) in improving travel behavior models. Also, within the *Transduction* step of the *Production-Transduction* framework is the recent study of the systematic heterogeneity in the disutility of travel time (Etezady et al., 2019), where an approach of transferring model coefficients rather than variables is investigated. In the future, a comparison of the two *Transduction* approaches could be of interest as well.

As the data becomes more available, the learning functions become more accurate, albeit also more complex. The complexity of such models buries the most ardent efforts to understand them under the billions of parameters involved, leaving the landscape dotted with “black boxes”. In many cases these learning functions take a lot of computation time, memory, and storage to train. This creates a fertile ground for experiments with pre-existing *Production* steps, in which refining only the most appropriate learning function to the given data is required.

If taken to the extreme degree, a substantial number of *Production-Transduction* exercises could create a multidimensional manifold of interdependency rules between any pair of variables, given the other variables, space, and time. Such a knowledge database would provide readily available data to address an immense number of research questions, for which data collection would be needed only to update the existing recorded interdependencies. Thus, *Production-Transduction*, besides immediate and narrowly defined benefits to various areas of knowledge (e.g., travel demand modeling in the case of this work), presents a greater benefit to social science related research and understanding.

APPENDIX A. INVESTIGATION OF JOINT CHOICE NESTED LOGIT MODELS

A.1 Issues Associated with Nested Logit Model Estimation

The quest for the best specification of the weighted nested logit model for the joint choice of mode and laptop usage (estimated with NLOGIT 6 and shown in Table A.1, with the pertinent attitudinal factors defined in Table A.2) revealed several issues with this approach. First, estimating the two choices jointly meant that data missing on any explanatory variable in any of the laptop usage models resulted in excluding the associated case from the entire estimation. This led to problems with further unbalancing shares that were already unbalanced. For example, in the stand-alone laptop/tablet usage models of Table 2.5, there are three bicycle choosers who used a laptop or tablet on their commute (see footnote 2 of the paper for an explanation), whereas in the nested logit model of Table A.1, that number has been reduced to zero[19]. Similarly, there are 37 laptop/tablet users among solo drivers in Table 2.5, but only 29 in Table A.1. We speculate that the difficulties we encountered in finding stable specifications of the nested logit model are probably in large measure a result of complete or quasicomplete separation problems arising from the unbalanced shares (Zorn, 2005).

---

[19] Even for the stand-alone binary choice model of Table 2.5, where there were three laptop/tablet choosers out of 268 bicyclists, we used only a constant term to specify the model. That is all the more necessary here, and even so, the constant is (1) large and negative, consistent with there being an essentially zero probability of choosing the associated alternative of laptop-with-bike; and (2) estimated with a large standard error, signifying substantial instability/imprecision (which is logical: there is a very wide range of parameter estimates which would yield an "essentially zero" probability of choice).

Second, multiple experiments showed that specifying the lower nest models (laptop usage choice) produces different estimates depending on which underlying utility function (for "use laptop" or "do not use laptop") an explanatory variable is associated with (an analytical proof of this is shown in Section A.2, for the special case in which each lower nest has two alternatives, and all variables are "assigned" to only one or the other of the two alternatives; the empirical evidence for one pair of specifications is presented in Table A.3 and Table A.4). This creates an ambiguity in how to specify each utility equation: for joint choice models with a binary secondary decision, there is little conceptual differentiation between the alternative specifications, which makes the model building process rather arbitrary. In particular, the estimates of the inclusive value (IV) parameter are also affected, and in the pair of models shown in Table A.3 and Table A.4, the different values obtained for that parameter could have led to different conclusions based on the statistical test for equivalency of the nested logit model to MNL (i.e., the t-test of the null hypothesis that the IV parameter is equal to one). In the joint estimation reported in Table A.1, for consistency we specified only "use laptop" lower nest functions, leaving their "do not use laptop" counterparts as the base ($V_{not\ used\ laptop,mode} = 0$). However, some experimentation showed that a marginal (possibly negligible[20]) improvement over the

[20] In one case, moving the single variable *Explorer* (*Driving Alone*-specific) from the laptop to no-laptop branch in the context of the best specification (Table A.1) changed the final log likelihood to –1207.262 (compared to –1207.575 for the base model reported in Table A.1; here and below the base model estimates are in parentheses). The estimated coefficient for that variable changed to 0.05250 (-0.05034). The inclusive value parameter, θ, for *Driving Alone* was 0.11596 (0.11162). The magnitudes for the upper nest estimated coefficients were also noticeably different between the two models (unlike the case in Table A.3 and Table A.4.)

current best specification (with respect to the goodness-of-fit measures) could be achieved by switching one or more of the explanatory variables to the "did not use laptop" function.

The third issue is that joint estimation, which applies some parameters such as weighting systematically across the levels, could lead to undesirable effects. The problem of weighting the current nested logit model formulation is two-fold: a) the true *joint* weights are unknown, and b) weighting adversely impacts the laptop usage nests for modes with a minor presence in the population. To elaborate: while the regional commute mode shares are known, this is not the case for laptop usage while commuting, let alone the joint distribution of commute mode and laptop usage. To develop our joint models, we circumvented this ignorance by (daringly) assuming that the mode-specific laptop usage shares observed in the sample are representative of the broad population. Setting the issue of unknown weights aside, applying the population mode shares to the lower nest is still undesirable. Since in the region driving alone occupies the lion's share of the commute (77%), the lower nest models are affected proportionally: the significance of the explanatory variables in the laptop usage nests is boosted for those modes with a weight higher than one (driving alone and shared ride), and is hampered for those modes with a weight lower than one (transit, biking, and commuter rail). Consequently, we blur a nuanced view of the explanatory-variables landscape of the travel-multitasking phenomenon, notably, areas of that landscape that are of special interest – those pertaining to collective modes.

Table A.1 – Weighted nested logit model for the joint choice of transportation mode and laptop usage

Variables	Biking	Commuter rail	Transit	Shared ride	Driving alone
TRANSPORTATION MODE CHOICE (upper nest)					
	Socioeconomic characteristics[a]				
Driver's license	–[b]	–	–2.043** (0.8291)	–	base
Female	–	–	–	0.345** (0.1557)	base
Race: white	–	–	0.601** (0.2352)	–	base
Limitation on walking	–	–	–	0.162*** (0.0599)	base
	Objective mode attributes				
In-vehicle travel time, min	–0.149** (0.0591)	· · ·	–0.013***c (0.0060)	· · ·	· · ·
Out-of-vehicle travel time, min	· · ·	· · ·	–0.046*** (0.0093)	· · ·	· · ·
One-way commute cost, ln($)	· · ·	· · ·	–1.189*** (0.1476)	· · ·	· · ·
	General attitudes				
Pro-active modes	2.104*** (0.4646)	–	–	–	base
Pro-transit	–	0.906*** (0.3369)	0.804*** (0.1236)	0.205** (0.0858)	base
	Multitasking preference				
Polychronicity	–	–	–	0.190** (0.0744)	base
	Mode perceptions				
Mode convenience	· · ·	· · ·	0.459*** (0.0668)	· · ·	· · ·
Mode benefit /cost	· · ·	· · ·	0.391*** (0.0723)	· · ·	· · ·
Mode comfort	· · ·	· · ·	0.423*** (0.0606)	· · ·	· · ·
Mode multitaskability	· · ·	· · ·	0.118** (0.0455)	· · ·	· · ·
	Constants				
Constant	–5.691*** (1.1365)	–3.286*** (0.4481)	0.853 (0.8314)	–2.648*** (0.2314)	base

Table A.1 (continued)

Variables	Biking	Commuter rail	Transit	Shared ride	Driving alone
LAPTOP USAGE CHOICE[d] (lower nest)					
General attitudes					
Pro-technology	–	–	0.199** (0.1012)	–	0.068*** (0.0187)
Travel is wasted time	–	–	–	0.102*** (0.0370)	
Pro-active modes	–	–	–	0.095*** (0.0366)	–0.071*** (0.0238)
Pro-transit	–	–	–	–	0.120*** (0.0334)
The only benefit of my job is money to do other things.	–	–	–	–	0.130*** (0.0257)
I'd be willing to give up a day's pay to get a day off.	–	–	–	–	–0.055** (0.0240)
I (would) like to own a car that impresses others.	–	–	–	–	0.048** (0.0207)
Personality traits					
Extraverted	–	–	–	–	0.046** (0.0229)
Leader	–	–	–	–	–0.053*** (0.0171)
Explorer	–	–	–	–	–0.050** (0.0209)
Multitasking preference					
Multitasking preference (polychronicity)	–	–	0.187** (0.0861)	–	–
Multitasking is normative	–	–	–	–	0.111*** (0.0275)
Time use					
Has to work on commute	–	0.879** (0.3416)	–	0.210*** (0.0387)	0.070*** (0.0236)
Has to multitask at work	–	–	–	–0.098*** (0.0366)	–
Would like to do recreation on commute	–	–	–	0.106*** (0.0346)	–0.132*** (0.0268)
Would like to take same route	–	–	–	–0.111*** (0.0382)	–0.080*** (0.0200)
Has to be available to people	–	–	–	0.114*** (0.0412)	–
Time spent working	–	–	–	–0.075*** (0.0273)	–
Time spent for ICT-enabled leisure and social activities[e]	–	–	–	0.074*** (0.273)	0.068*** (0.0145)
Has to do recreation on commute	–	–	–	–	0.207*** (0.0393)

Table A.1 (continued)

Variables	Biking	Commuter rail	Transit	Shared ride	Driving alone
Would like to work on commute	–	–	–	–	0.063*** (0.0227)
Time spent for non-ICT-enabled leisure and social activities	–	–	–	–	–0.066*** (0.0145)
			Attitudes toward waiting[f]		
Don't mind waiting	–	–	–	–	0.114*** (0.0290)
Don't need to be equipped for a wait event	–	–	–	–	–0.051** (0.0250)
			Socioeconomic characteristics		
Annual household per capita income, $000	–	–	–	0.005*** (0.0012)	–
Travel distance, mi	–	–	–	0.006*** (0.0016)	–
Race: white	–	–	–	–0.228*** (0.0781)	–
Vehicle age	–	–	–	–	–0.019*** (0.0058)
Vehicle availability[g]	–	–	–	–	–0.515*** (0.1264)
Occupation: service	–	–	–	–	0.354*** (0.0994)
Share of time vehicle is available	–	–	–	–	–0.065*** (0.0185)
Race: black	–	–	–	–	0.246*** (0.0661)
			Constants		
Constant	–6.548 (82.0408)	–0.819 (0.6498)	–0.757*** (0.1014)	–0.652*** (0.1020)	0.384** (0.1630)
		INCLUSIVE VALUE PARAMETERS			
θ	0.151 (1.8922)	0.862*** (0.0237)	0.337*** (0.0394)	0.125*** (0.0160)	0.112*** (0.0109)
SD of ε	0.194 (2.4268)	1.105*** (0.0303)	0.432*** (0.0506)	0.160*** (0.0205)	0.143*** (0.0139)

Number of observations	1948		
$\mathcal{L}(\mathbf{0})$	–4145.429	$-2(\mathcal{L}(\mathbf{0}) - \mathcal{L}(\hat{\boldsymbol{\beta}}))$	5875.708
$\mathcal{L}(\boldsymbol{c})$	–1799.691	ρ^2	0.7087
$\mathcal{L}(\hat{\boldsymbol{\beta}})$	–1207.575	Adjusted ρ^2	0.6918

***, ** = significant at 1%, 5%.

[a] Effects of the variables are represented by an estimated coefficient and standard error (in parentheses).

[b] Dashes indicate coefficients that were constrained to be zero after they were found to have significance > 0.05.

[c] Centered coefficients with no dashes in the row represent generic coefficients (equal across alternatives).

[d] Explanatory variables were specified only in the "used-laptop" functions in each nest. The "not-used-laptop" functions were the base (i.e., = 0).

[e] Standardized responses for this single item.

[f] Constructs are defined in Mishra et al. (2015).

[g] A ratio between household vehicles and licensed drivers, capped at 1.

Table A.2 – Personality traits and time use constructs used in the joint choice model of Table A.1

Constructs	Statements[a]	Loadings[b]
	Personality traits[c]	
Extraverted	Fun-oriented	0.694
	Spontaneous	0.601
	Variety-seeking	0.537
	Adventurous	0.520
	Like to meet new people	0.439
	Risk-taking	0.308
Leader	Ambitious	0.698
	Work-oriented	0.513
	Like being in charge	0.373
	Efficient	0.318
Explorer	Concerned about the environment	0.751
	Curious	0.494
	Like being outdoors	0.396
	Time use[d]	
Time spent for non-ICT-enabled leisure and social activities	With friends	0.585
	Doing hobbies	0.427
	Getting exercise	0.379
	With family	0.369
	Volunteering/ doing service	0.320

[a] A statement can load on more than one construct.
[b] Represents the degree of association between the statement and the construct. Only loadings greater than 0.3 in magnitude are reported.
[c] How well the item describes the respondent is measured on a 5-point Likert-type scale ranging from "Hardly at all" to "Almost completely".
[d] Items measured on a 5-point ordinal scale ranging from "Way too little" to "Way too much".

Finally, the joint estimation poses additional challenges for constructing "what-if" scenarios (Section 2.7), which may help in assessing the aggregate effect of travel multitasking on current and future regional commute mode shares. For these reasons, we adopted the alternative approach presented in the main body of the paper.

A.2 Proof that Coefficients in a Nested Logit Model Differ When Variables in (Binary Choice) Lower Nests are Associated with Different Alternatives

Let m index the modes in the upper level, m = bicycle (B), rail (R), transit (T), shared ride (S), and drive alone (D).

Let ℓ index the two alternatives in the lower nests, ℓ = laptop used (L), not used (N). In the following discussion, we suppress the subscript denoting the individual case, for simplicity.

Let $V_m = \beta_m' x_m$ be the systematic portion of utility pertaining to mode m (without the inclusive value term from the lower nest), and let $V_\ell = \gamma_\ell' z_\ell$ be the systematic portion of utility pertaining to nest alternative ℓ, where x and z are vectors of explanatory variables, and β and γ the respectively associated vectors of coefficients.

If we had only an unnested binary choice between L and N, it would not matter "where the z's went": only differences in utility matter (Train, 2009, p. 19), so for the binary logit model we would have (where θ is the scale parameter of the extreme value distribution):

$$\begin{aligned} P_L &= \frac{e^{V_L/\theta}}{e^{V_L/\theta}+e^{V_N/\theta}} = \frac{e^{\gamma' z_L/\theta}}{e^{\gamma' z_L/\theta}+e^{\gamma' z_N/\theta}} \\ &= \frac{1}{1+e^{-\gamma'(z_L-z_N)/\theta}} = \frac{1}{1+e^{\gamma'(z_N-z_L)/\theta}}, \end{aligned} \tag{3}$$

a function only of $(z_L - z_N)$, and

$$P_N = 1 - P_L = \frac{e^{V_N/\theta}}{e^{V_L/\theta} + e^{V_N/\theta}} = \frac{e^{\gamma' z_N/\theta}}{e^{\gamma' z_L/\theta} + e^{\gamma' z_N/\theta}}$$
$$= \frac{1}{1 + e^{-\gamma'(z_N - z_L)/\theta}} = \frac{1}{1 + e^{\gamma'(z_L - z_N)/\theta}}. \tag{4}$$

For simplicity and without loss of generality, suppose that $z_L' z_N = 0$, i.e. that if the k^{th} element of z_L is *not* zero then the corresponding k^{th} element of z_N *is* zero, and similarly for z_N. That is, all explanatory variables in this binary choice model are "assigned" to one alternative or the other (not necessarily all to the *same* alternative), and take on the value 0 for the opposite alternative.

For ease of exposition, we will consider the special case in which all explanatory variables *are* assigned to one alternative or the other – the resulting pair of models could be termed the "stereoisomers" (or "enantiomers") of binary discrete choice modeling. We also assume (as is the case in our application) that all variables in the binary choice model are individual-specific rather than alternative-specific, so that there is no reason for the sign of a *variable* (as opposed to the *coefficient*) to change if it is assigned to one alternative rather than the other[21], in which case we can refer to the single vector of explanatory

[21] On the other hand, if a given variable differed by alternative, for example if the lower level involved access modes, each with its own travel time and cost, then it would be natural for the difference to change sign if it were assigned to the "other" alternative: $(z_L - z_N)$ would become $(z_N - z_L) = -(z_L - z_N)$ if the base alternative changed from N to L (if the travel time on mode L took 10 minutes *longer* than on mode N, $TT_L - TT_N = 10$, then the travel time on N is 10 minutes *less* than on L: $TT_N - TT_L = -10$). If all variables were like this, then from eq.(3) we would simply have $P_{L,Lnon-zero} = \frac{1}{1 + e^{-\gamma_L'(z_L - z_N)/\theta}}$ when N is the base alternative and

variables as *z*. Let γ_L be the vector of parameters that results when z_L is non-zero (i.e., = *z*) and z_N is zero (in this case, N is technically the base alternative, but to reduce confusion we will refer to this case as "Lnon-zero" and use L subscripts to distinguish it from the other case), and γ_N be the vector of parameters that results when z_N is non-zero (i.e., = *z*) and z_L is zero ("Nnon-zero").

Then under this common set of circumstances, when z_L is non-zero and z_N is zero, from eq. (3) we have:

$$P_{L,Lnon-zero} = \frac{1}{1+e^{-\gamma_L' z/\theta}},$$

and when z_N is non-zero and z_L is zero we have:

$$P_{L,Nnon-zero} = \frac{1}{1+e^{\gamma_N' z/\theta}}.$$

Equality of these two probabilities implies that $\gamma_L/\theta = -\gamma_N/\theta$, as would be expected. In other words, in a simple binary choice model, if all the variables are "switched" from one alternative to the other without changing signs, then the estimated coefficients will differ only in that their signs will be reversed.

$P_{L,Nnon-zero} = \frac{1}{1+e^{\gamma_N'(z_N - z_L)/\theta}} = \frac{1}{1+e^{-\gamma_N'(z_L - z_N)/\theta}}$ when L is the base alternative, which means that for the two probabilities to be equal, the coefficients would be equal: $\gamma_L/\theta = \gamma_N/\theta$. If the z variables are a *mixture* of individual-specific and alternative-varying, then with a change in the base, the signs of the corresponding γ coefficients would reverse (while the variable would remain the same) or remain the same (while the variable difference would reverse), respectively.

We turn now to the nested logit model. For simplicity of exposition, we will maintain the special case in which all explanatory variables in the lower nests are individual-specific and all are assigned to the same alternative, whichever that may be. We will demonstrate that switching all variables from one alternative to the other will not result in a simple reversal of signs. This result can be generalized to the more complex case in which only some variables are switched, but the demonstration of the simpler case will suffice for our purposes.

In the nested logit model, for the choice of laptop given mode *m* we have

$$P_{L|m} = \frac{e^{V_L/\theta_m}}{e^{V_L/\theta_m} + e^{V_N/\theta_m}} = \frac{1}{1+e^{-\gamma'(z_L - z_N)/\theta_m}}, m = B, R, T, S, D, \quad (5)$$

and for the choice of mode *m* we have

$$P_m = \frac{e^{V_m + \theta_m \Gamma_m}}{\sum_{m'} e^{V_{m'} + \theta_{m'} \Gamma_{m'}}}, \quad (6)$$

where

$$\Gamma_m = \ln\left[e^{V_L/\theta_m} + e^{V_N/\theta_m}\right] = \ln\left[e^{\gamma' z_L/\theta_m} + e^{\gamma' z_N/\theta_m}\right].$$

Then for the Lnon-zero case we will have $z_L = z$, $z_N = 0$, and, from eq. (5),

$$P_{L|m,Lnon-zero} = \frac{1}{1+e^{-\gamma_L' z/\theta_{mL}}},$$

while for the Nnon-zero case we will have $z_L = 0$, $z_N = z$, and, again from eq. (5),

$$P_{L|m,Nnon-zero} = \frac{1}{1+e^{\gamma_N ' z/\theta_{mN}}}.$$

Equality of these two probabilities implies that $-\gamma_L ' z / \theta_{mL} = \gamma_N ' z / \theta_{mN}$, or, on an element-by-element basis of the coefficient vectors, that $-\gamma_L / \theta_{mL} = \gamma_N / \theta_{mN}$.

At the same time, for the Lnon-zero case we have

$$\Gamma_{m,Lnon-zero} = \ln\left[e^{\gamma_L ' z/\theta_{mL}} + e^0\right] = \ln\left[1 + e^{\gamma_L ' z/\theta_{mL}}\right] \tag{7}$$

while for the Nnon-zero case we have

$$\Gamma_{m,Nnon-zero} = \ln\left[e^0 + e^{\gamma_N ' z/\theta_{mN}}\right] = \ln\left[1 + e^{\gamma_N ' z/\theta_{mN}}\right] \tag{8}$$

We can express $\Gamma_{m,Nnon-zero}$ as a function of $\Gamma_{m,Lnon-zero}$ by replacing γ_N / θ_{mN} with $-\gamma_L / \theta_{mL}$ in eq. (8):

$$\begin{aligned}
\Gamma_{m,Nnon-zero} &= \ln\left[1 + e^{\gamma_N ' z/\theta_{mN}}\right] = \ln\left[1 + e^{-\gamma_L ' z/\theta_{mL}}\right] \\
&= \ln\left[\left(\frac{e^{\gamma_L ' z/\theta_{mL}}}{e^{\gamma_L ' z/\theta_{mL}}}\right)\left(1 + e^{-\gamma_L ' z/\theta_{mL}}\right)\right] = \ln\left[e^{-\gamma_L ' z/\theta_{mL}}\left(e^{\gamma_L ' z/\theta_{mL}} + 1\right)\right] \\
&= \ln\left(e^{-\gamma_L ' z/\theta_{mL}}\right) + \ln\left[e^{\gamma_L ' z/\theta_{mL}} + 1\right] = -\gamma_L ' z/\theta_{mL} + \Gamma_{m,Lnon-zero}.
\end{aligned} \tag{9}$$

Thus, $\Gamma_{m,Nnon-zero} \neq \Gamma_{m,Lnon-zero}$; the difference is the term $-\gamma_L' z / \theta_{mL}$. Note that because the Γ_m s are different, the θ_m multiplying the Γ_m in eq. (6) will (in general) have a different estimated value, which means that the γ / θ_m coefficients of eq. (5) will also have different estimated values. In other words, the coefficients of the models at both levels will be affected.

This analytically-derived difference is confirmed empirically for the illustrative pair of models shown in Table A.3 and Table A.4, in which the bicycle mode has been removed and the lower level models are specified only with constant terms. Note that in this case, the final log-likelihoods are identical between the two models, and the parameters of the upper-level model differ only in the constant terms (although the standard errors of all parameters differ somewhat). This is as expected, since when the lower level is specified only with constants, $-\gamma_L' z / \theta_{mL}$ is just the constant $-\gamma_L / \theta_{mL}$, from eq. (7) $\Gamma_{m,Lnon-zero}$ is also a constant, and in eq. (6), $\theta_{mL}\Gamma_{m,Lnon-zero}$ is also a constant – and similarly for the Nnon-zero case (see the bottom block of Table A.4, in which the quantity $-\gamma_L' z / \theta_{mL}$ is shown to equal the difference in Γs, per eq. (9)). The differing values of these constant inclusive value terms between the two cases will simply shift the constant terms of the upper-level model to compensate. Letting $\beta_{0m,L}$ and $\beta_{0m,N}$ be the constant term for mode *m* corresponding to the Lnon-zero and Nnon-zero cases respectively, it can be confirmed from Table A.3 and Table A.4 that (within round-off error)

$$\beta_{0m,L} + \theta_{mL}\Gamma_{m,Lnon-zero} = \beta_{0m,N} + \theta_{mN}\Gamma_{m,Nnon-zero}, \tag{10}$$

after *subsequently* shifting all constants in each model to set the last mode (driving alone) as the base (whose constant is fixed at 0) as is done by default in NLOGIT.

However, as mentioned in Section A.1, the two estimates of the inclusive value parameter (together with their standard errors) could, depending on the conservatism of the analysis, yield different conclusions about whether the nested logit model were statistically equivalent to MNL. For the model of Table A.3, the t-statistic of the test is (0.982 – 1) / 0.0083 = –2.17 ($p = 0.03$), with the point estimate of 0.982 implying an error correlation (between alternatives in the nest) of $(1 - 0.982^2) = 0.04$, which for all practical purposes indicates independence of the errors. For the model of Table A.4, by contrast, the t-statistic is (0.783 – 1) / 0.0502 = –4.32 ($p = 0.000016$), with the point estimate of 0.783 implying a (substantial) error correlation of 0.39.

Table A.3 – Weighted nested logit model for the joint choice of transportation mode and laptop usage, with the lower nest variables associated with "not-used-laptop" branch (*Nnon-zero*)

Variables	Commuter rail	Transit	Shared ride	Driving alone
TRANSPORTATION MODE CHOICE (upper nest)				
	Socioeconomic characteristics[a]			
Driver's license	–[b]	–2.349** (1.0126)	–	base
Female	–	–	0.280* (0.1513)	base
Race: white	–	0.600** (0.2383)	–	base
Limitation on walking	–	–	0.158*** (0.0590)	base
	Objective mode attributes			
In-vehicle travel time, min	· · · · · · · · · ·	· · · · · · · ·	–0.016^{***c} (0.0058)	· · · · · · · · · ·
Out-of-vehicle travel time, min	· · · · · · · · · ·	· · · · · · · ·	–0.049*** (0.0097)	· · · · · · · · · ·
One-way commute cost, ln($)	· · · · · · · · · ·	· · · · · · · ·	–1.419*** (0.1747)	· · · · · · · · · ·
	General attitudes			
Pro-transit	0.956*** (0.3166)	0.861*** (0.1256)	0.264*** (0.0842)	base
	Multitasking preference			
Polychronicity	–	–	0.215*** (0.0717)	base
	Mode perceptions			
Mode convenience	· · · · · · · · · ·	· · · · · · · ·	·0.418*** (0.0662)	· · · · · · · · · ·
Mode benefit /cost	· · · · · · · · · ·	· · · · · · · ·	·0.348*** (0.0759)	· · · · · · · · · ·
Mode comfort	· · · · · · · · · ·	· · · · · · · ·	·0.413*** (0.0620)	· · · · · · · · · ·
Mode multitasking	· · · · · · · · · ·	· · · · · · · ·	·0.110** (0.0462)	· · · · · · · · · ·
	Constants			
Constant	–0.285 (0.5014)	2.119** (1.0426)	–1.174*** (0.3030)	base

Table A.3 (continued)

Variables	Commuter rail	Transit	Shared ride	Driving alone
LAPTOP USAGE CHOICE[d] (lower nest)				
	Constants			
Constant	0.102 (0.5159)	2.211*** (0.2589)	1.451*** (0.1582)	3.210** (0.1271)
INCLUSIVE VALUE PARAMETERS				
θ	· · · · · · · · · · · · · · · ·	· · ·	·0.982*** (0.0083)	· · · · · · · · · ·
SD of ε	· · · · · · · · · · · · · · · ·	· · ·	·1.259*** (0.0107)	· · · · · · · · · ·
Number of observations	2010	$\mathcal{L}(\hat{\boldsymbol{\beta}})$		–1404.899
$\mathcal{L}(\mathbf{0})$	–3691.503	ρ^2		0.6194

***, **, * = significant at 1%, 5%, and 10%.

[a] Effects of the variables are represented by an estimated coefficient and standard error (in parentheses).

[b] Dashes indicate coefficients that were constrained to be zero after they were found to have significance > 0.05.

[c] Centered coefficients with dots across the row represent generic parameters (constrained to be equal across alternatives).

[d] Explanatory variables were specified only in the “not used-laptop” functions in each nest. The “used-laptop” functions were the base (i.e., = 0).

Table A.4 – Weighted nested logit model for the joint choice of transportation mode and laptop usage with the lower nest variables associated with "used-laptop" branch (*Lnon-zero*)

Variables	Commuter rail	Transit	Shared ride	Driving alone
TRANSPORTATION MODE CHOICE (upper nest)				
	Socioeconomic characteristics[a]			
Driver's license	–[b]	–2.349** (1.0189)	–	base
Female	–	–	0.280* (0.1518)	base
Race: white	–	0.600** (0.2406)	–	base
Limitation on walking	–	–	0.158*** (0.0594)	base
	Objective mode attributes			
In-vehicle travel time, min	· · · · · · · · · · · · · · –0.016***[c] (0.0059) · · · · · · · · · · ·			
Out-of-vehicle travel time, min	· · · · · · · · · · · · · · –0.049*** (0.0110) · · · · · · · · · · ·			
One-way commute cost, ln($)	· · · · · · · · · · · · · · –1.419*** (0.2114) · · · · · · · · · · ·			
	General attitudes			
Pro-transit	0.956*** (0.3354)	0.861*** (0.1310)	0.264*** (0.0892)	base
	Multitasking preference			
Polychronicity	–	–	0.215*** (0.0725)	base
	Mode perceptions			
Mode convenience	· · · · · · · · · · · · · · 0.418*** (0.0775) · · · · · · · · · · ·			
Mode benefit /cost	· · · · · · · · · · · · · · 0.348*** (0.0782) · · · · · · · · · · ·			
Mode comfort	· · · · · · · · · · · · · · 0.413*** (0.0622) · · · · · · · · · · ·			
Mode multitasking	· · · · · · · · · · · · · · 0.110** (0.0462) · · · · · · · · · · ·			
	Constants			
Constant	–3.273*** (0.5290)	1.132 (1.0168)	–2.900*** (0.3267)	base
LAPTOP USAGE CHOICE[d] **(lower nest)**				
	Constants			
Constant	–0.082 (0.4501)	–1.764*** (0.2327)	–1.158*** (0.1325)	–2.561** (0.1185)

Table A.4 (continued)

Variables	Commuter rail	Transit	Shared ride	Driving alone
INCLUSIVE VALUE PARAMETERS				
θ	· · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · ·0.783***	· · · · · · · · · ·	· · · · · · · · · ·
		(0.0502)		
SD of ε	· · · · · · · · · · · · · · ·	· · · · · · · · · · · · · · ·1.005***	· · · · · · · · · ·	· · · · · · · · · ·
		(0.0643)		
QUANTITIES TRANSLATING BETWEEN *Lnon-zero* and *Nnon-zero* MODELS				
$\Gamma_{m,Lnon-zero}$ (eq. (5))	0.642	0.100	0.205	0.037
$\Gamma_{m,Nnon-zero}$ (eq. (6))	0.747	2.352	1.684	3.307
$-\gamma_L ' z / \theta_{mL}$	0.105	2.252	1.479	3.270
$\Gamma_{m,Nnon-zero} - \Gamma_{m,Lnon-zero}$	0.105	2.252	1.479	3.270
unshifted $\beta_{0m,L} + \theta_{mL}\Gamma_{m,Lnon-zero}$	–2.770	1.211	–2.739	0.029
shifted $\beta_{0m,L} + \theta_{mL}\Gamma_{m,Lnon-zero}$	–2.799	1.181	–2.768	0.000
unshifted $\beta_{0m,N} + \theta_{mN}\Gamma_{m,Nnon-zero}$	0.448	4.428	0.479	3.247
shifted $\beta_{0m,N} + \theta_{mN}\Gamma_{m,Nnon-zero}$	–2.799	1.181	–2.768	0.000
Number of observations	2010	$\mathcal{L}(\hat{\boldsymbol{\beta}})$		–1404.899
$\mathcal{L}(\mathbf{0})$	–3691.503	ρ^2		0.6194

***, **, * = significant at 1%, 5%, and 10%.
[a] Effects of the variables are represented by an estimated coefficient and standard error (in parentheses).
[b] Dashes indicate coefficients that were constrained to be zero after they were found to have significance > 0.05.
[c] Centered coefficients with dots across the row represent generic parameters (constrained to be equal across alternatives).
[d] Explanatory variables were specified only in the "used-laptop" functions in each nest. The "not-used-laptop" functions were the base (i.e., = 0).

A.3 Model Equations

Following random utility theory, in a discrete choice problem an individual will choose an alternative that maximizes his or her utility. The utility U that each individual n associates with an alternative i can be decomposed into the deterministic V and stochastic ε parts:

$$U_{i,n} = V_{i,n} + \varepsilon_{i,n} \ .$$

Assuming that the error term has the extreme-value distribution, i.e., $\varepsilon \sim EV(0, \mu)$, the probability of choosing the i^{th} alternative is given by:

$$P_n(i) = \Pr\left(U_{i,n} \geq U_{j,n}\ \forall j \in J_n\right) = \frac{e^{\mu V_{i,n}}}{\sum_{j=i}^{J_n} e^{\mu V_{j,n}}} ,$$

where j is another alternative from the individual's choice set J_n and μ is the scaling parameter, which conventionally is set to unity.

The rest of this section presents the deterministic parts of the utility equations estimated in Sections 2.5 and 2.6 for the five binary logit models of mode-specific propensity to use a laptop, netbook, or tablet computer (Table 2.5) and the multinomial logit commute mode choice model (Table 2.6), respectively. In these equations the variable names have been abridged for greater readability. Table A.5 provides the correspondence between the abbreviations in the equations and the variable names used in the rest of the paper.

Table A.5 – Dictionary for model equations

Abbreviation of variable name in equations	Variable name
Has2workOnCommute	Has to work on commute
Like2takeSameRoute	Would like to take same route
isFemale	Female
Age	Age
isHourlyWaged	Hourly waged
TravelDistance	Travel distance
isProTech	Pro-technology
isPolychronic	Multitasking preference (polychronicity)
ThinksTravelIsTimeWasted	Travel is wasted time
Likes2recOnCommute	Would like to do recreation on commute
Has2MTatWork	Has to multitask at work
Likes2beAvailable	Would like to be available to people
HHIncomePerCapita	Annual household per capita income
ThinksMTisNormative	Multitasking is normative
TimeSpentWorking	Time spent working
Has2recOnCommute	Has to do recreation on commute
VehicleAge	Vehicle age
IVTT	In-vehicle travel time
OVTT	Out-of-vehicle travel time
Cost	One-way commute cost
ModeConvinience	Mode convenience
ModeBenefitCost	Mode benefit /cost
ModeComfort	Mode comfort
ModeMultitaskability	Mode multitaskability
Propens2useLaptop	Propensity to use a laptop/ tablet/ netbook
isProTransit	Pro-transit
hasDriverLic	Driver’s license
isWhite	Race: white
isProActiveModes	Pro-active modes
LimitsWalking	Limitation on walking

A.1.1 Binary logit models of mode-specific propensity to use a laptop, netbook, or tablet computer (Table 2.5)

$V_{biking} = -4.470$

$V_{commuterRail} = -0.313 + 1.148 * Has2workOnCommute - 0.543 * Likes2takeSameRoute - 1.36 * isFemale - 0.049 * Age - 3.276 * isHourlyWaged + 0.026 * TravelDistance$

$V_{transit} = -2.268 + 0.549 * isProTech + 0.241 * isPolychronic + 0.368 * Has2workOnCommute$

$V_{sharedRide} = -4.408 + 0.564 * ThinksTravelIsTimeWasted + 1.262 * Has2workOnCommute + 0.685 * Likes2recOnCommute - 0.456 * Has2MTatWork + 0.486 * Likes2beAvailable - 0.383 * Likes2takeSameRoute - 0.021 * HHIncomePerCapita + 0.029 * TravelDistance$

$V_{drivingAlone} = -2.178 + 0.401 * ThinksMTisNormative - 0.372 * TimeSpentWorking + 0.77 * Has2workOnCommute + 0.946 * Has2recOnCommute - 0.389 * Likes2recOnCommute - 0.102 * VehicleAge$

A.1.2 Multinomial logit commute mode choice model (Table 2.6)

$V_{biking} = -5.327 - 0.163 * IVTT - 0.048 * OVTT - 1.175 * \ln(Cost) + 2.088 * isProActiveModes + 0.455 * ModeConvinience + 0.368 * ModeBenefitCost + 0.405 * ModeComfort + 0.098 * ModeMultitaskability + 1.240 * Propens2useLaptop$

$$V_{commuterRail} = -2.959 - 0.016 * IVTT - 0.048 * OVTT - 1.175 * \ln(Cost) + 0.954 * isProTransit + 0.455 * ModeConvinience + 0.368 * ModeBenefitCost + 0.405 * ModeComfort + 0.098 * ModeMultitaskability + 1.240 * Propens2useLaptop$$

$$V_{transit} = 0.785 - 1.890 * hasDriverLic + 0.523 * isWhite - 0.016 * IVTT - 0.048 * OVTT - 1.175 * \ln(Cost) + 0.825 * isProTransit + 0.455 * ModeConvinience + 0.368 * ModeBenefitCost + 0.405 * ModeComfort + 0.098 * ModeMultitaskability + 1.240 * Propens2useLaptop$$

$$V_{sharedRide} = -2.752 + 0.393 * isFemale + 0.166 * LimitsWalking - 0.016 * IVTT - 0.048 * OVTT - 1.175 * \ln(Cost) + 0.201 * isProTransit + 0.191 * isPolychronic + 0.455 * ModeConvinience + 0.368 * ModeBenefitCost + 0.405 * ModeComfort + 0.098 * ModeMultitaskability + 1.240 * Propens2useLaptop$$

$$V_{sharedRide} = -0.016 * IVTT - 0.048 * OVTT - 1.175 * \ln(Cost) + 0.455 * ModeConvinience + 0.368 * ModeBenefitCost + 0.405 * ModeComfort + 0.098 * ModeMultitaskability + 1.240 * Propens2useLaptop$$

APPENDIX B. DISTRIBUTIONS OF VALUE OF TRAVEL TIME AND WILLINGNESS TO PAY TO USE A LAPTOP

Table B.1 – Distribution of VOTT and WTP to use laptop in millennial segment

	Mean	**5th%**	**25th%**	**Median**	**75th%**	**95th%**
WTP Rail-DA, $	1.0628	–0.2712	–0.0029	0.0312	0.6416	6.3360
WTP Rail-Transit, $	0.6973	–0.9289	–0.1662	–0.0005	0.2967	4.9913
WTP Rail-DA, min	17.1155	–11.1260	–0.3396	3.1704	23.5844	98.6188
WTP Rail-Transit, min	7.4873	–28.0932	–8.0282	–0.2191	10.8901	81.9448
Value of IVTT, $/hr	1.8971	0.1171	0.3426	1.4233	2.2559	6.9150
Value of OVTT, $/hr	5.7401	0.0249	1.0466	4.3487	6.8928	21.1282

Table B.2 – Distribution of VOTT and WTP to use laptop in non-millennial segment

	Mean	**5th%**	**25th%**	**Median**	**75th%**	**95th%**
WTP Rail-DA, $	0.4047	–0.0949	–0.0049	0.0200	0.2003	2.0419
WTP Rail-Transit, $	0.2566	–0.3047	–0.0874	–0.0155	0.0819	1.5790
WTP Rail-DA, min	5.3914	–3.0566	–0.2778	1.1032	7.0226	30.5848
WTP Rail-Transit, min	1.7840	–7.9845	–3.5008	–1.0723	3.0008	25.6362
Value of IVTT, $/hr	2.3101	0.0593	0.9427	1.7961	2.5052	7.5916
Value of OVTT, $/hr	7.0928	0.0253	2.8983	5.5223	7.7025	23.3409

Table B.3 – Distribution of VOTT and WTP to use laptop in the whole sample

	Mean	**5th%**	**25th%**	**Median**	**75th%**	**95th%**
WTP Rail-DA, $	0.6087	–0.1454	–0.0059	0.0282	0.3189	3.0987
WTP Rail-Transit, $	0.3890	–0.4732	–0.1241	–0.0145	0.1394	2.5245
WTP Rail-DA, min	8.2697	–4.5289	–0.3728	1.7006	10.7562	47.9364
WTP Rail-Transit, min	2.9697	–12.2085	–4.9533	–1.2111	4.9441	38.6440
Value of IVTT, $/hr	2.2763	0.0829	0.8058	1.7320	2.5669	7.4849
Value of OVTT, $/hr	6.7616	0.0251	2.4005	5.1595	7.6465	22.2966

APPENDIX C. TRANSFER LEARNING: ADDITIONAL BACKGROUND AND EXTENDED RESULTS

C.1 Additional Background and Review of Related Literature

C.1.1 Statistical Matching

The objective of this work could be achieved by *statistical matching* rather than *record linkage* and *multisensor data fusion*, so it is of interest to review particular methods that are implemented for statistical matching processes in the literature. To this end, D'Orazio et al. (2006) distinguish two approaches to data integration: macro and micro. In the macro approach, only the joint distribution of variables of interest (observed in one dataset and inferred in another) is transferred across the data sources. In the micro approach, the goal is to ascribe individual values to the variables of interest by inferring them via some approximation. The micro approach is most broadly used (including by the present study) due to the wider range of applications available with the resultant data. In a general case that involves two datasets, each of which consists of unique and common variables, the goal of the micro approach to statistical matching is to ascribe missing values of the unique variables in a combined stacked dataset using their partially observed relationship with the common variables.

However, there could be infinite ways of ascribing missing values to the unique variables of both datasets because they have been never observed jointly, thus, an identifiability problem exists. To overcome this problem, it is customary (and often implicit) to assume conditional independence between ascription targets, that is, given the common variables, the joint distributions of the unique variables are independent (D'Orazio et al., 2006). Viability of this assumption is extremely context-dependent; and in many real-world scenarios it is not guaranteed. In practice, there are two ways to relax the *conditional independence assumption* (CIA): (1) by expanding the common variables set, and (2) by collecting additional (small-batch) *auxiliary* data that observes all sets of unique and a set of common variables jointly (Fosdick et al., 2016; Schiefeling et al., 2016).

Typically, micro statistical matching consists of several steps. First, a model is trained on the fully observed data ("*donor*" or "*source*"), using unique and common portions as dependent and independent variables, respectively. Next, the trained model is applied to other datasets ("*recipients*" or "*targets*"), for which only the common variables are known. Finally, the predictions of this model are ascribed to the recipient datasets, synthetically supplying them with the previously unobserved variables. The result of all combinations between donors and recipients could be stacked to produce a complete synthetic dataset with unique variables defined across all observations. Different methods of micro statistical matching modify this algorithm to accommodate certain contexts and to improve validity.

Multiple factors are considered to classify methods of micro statistical matching: (1) type of function used (parametric, non-parametric, mixed, or Bayesian); (2) presence or absence of auxiliary data; (3) matching constraints used (e.g., "can an observed value be ascribed to multiple observations?" or "can a recipient observation with multiple missing values have different donor observations?"; Rässler, 2002); (4) levels of validity targeted (preserving individual values, joint distributions, correlation structures, and marginal distributions; Rässler, 2002); and (5) presence or absence of multiple outcome aggregation (in the case of multiple imputation).

Specifically, parametric methods rely on conditional mean matching (regression and log-linear for continuous and discrete ascription, respectively) to capture the observed relationship between unique (dependent) and common (independent) variables. Stochastic noise could be used to create additional variability in the ascribed values. Non-parametric methods do not estimate parameters of the matching function explicitly; rather, they learn the marginal and joint distributions of the variables in the training data implicitly. For example, the *random hot deck* (*RHD*) method ascribes values in the recipient dataset with random draws from the values of unique variables observed in the donor dataset (Andridge and Little, 2010). Variations of RHD include methods such as ranked hot deck and distance hot deck (D'Orazio et al., 2006).

While RHD uses the whole training dataset to predict values of the unique variables in the recipient dataset, the prediction accuracy could be improved if the consideration pool were limited only to similar observations. The *k-nearest neighbors* (*kNN*) method is a non-parametric, locally-approximated algorithm that implements this "informational" homogeneity. It works well with both continuous and discrete dependent variables. In the method, observations from both datasets are mapped in the hyperspace defined by the common variables. Then, for each observation from the recipient dataset, the *k* closest neighbors from the donor dataset are "polled", and the distribution of their "votes" (namely, the most-commonly-appearing class for predicting a categorical dependent variable, and an averaged value for predicting a continuous one) defines the ascribed value. The proximity of neighbors is determined by a Euclidean, weighted, or other distance function.

The value of *k* has an inverse relation with the complexity of the model and homogeneity of neighborhoods: larger *k*s correspond to fewer, more heterogeneous neighborhoods.

Hot deck and *kNN* methods and their close relatives were among the first to be implemented in the early history of statistical matching due to their simplicity and low computational complexity. While they are still very popular today because of requiring fewer assumptions about the data, more elaborate non-parametric methods are being proposed: for example, the Gibbs sampler approach (Ahfock et al., 2016) performs a search in a complex multi-dimensional restricted set to fill in values.

Mixed methods employ two-stage processes that include both parametric and non-parametric methods, which first approximate some value and then ascribe an observed value based on this approximation (D'Orazio et al., 2006). Finally, alternatively to the frequentist approach of modeling, *Bayesian* methods incorporate substantial randomness into the parametric approach by allowing parameter uncertainty and outcome noise to be determined by posterior probabilities observed from the data (Rässler, 2002; van Buuren, 2012). Some interesting recent examples of the approach include the Guided Bayesian Adjustment for Confounding framework that incorporate dimension reduction and treatment for heterogeneity (Antonelli et al., 2017).

The introduction of parameter uncertainty in Bayesian methods aligns rather well with the *multiple imputation* framework. *Imputation*, or treatment of statistical matching as a nonresponse phenomenon (Rässler, 2002), is an alternative perspective on the problem. As such, donor and recipient datasets could be concatenated by (column-wise) aligning common variables, and assigning missing values to the unique variables of observations from the recipient datasets. Afterward, a desired imputation method could be applied to "recover" the missing values. Multiple imputation for statistical matching was first proposed by Rubin (1986) as a way to overcome the CIA assumption and preserve the inherent uncertainty about true values of the unobserved unique variables in a recipient dataset. By using random parameter distribution draws, multiple imputation with chained equations creates several (*m*) datasets that show variability in the filled-in missing values but retains respective marginal and joint distributions across the concatenated datasets. An analyst, then, needs to average distributional parameters (e.g., mean, standard deviation, regression coefficients, etc.) across the datasets to arrive at unbiased (under missing completely at random and missing at random conditions) estimators. One apparent drawback of the method is the added analysis complexity of carrying along all imputed datasets and finding an average of *m* analyses. Attempting to overcome this drawback by the tempting shortcut of averaging imputed values across datasets and then proceeding with the single averaged dataset is not recommended because "imputation is not prediction" (van Buuren, 2012, p. 45). That is, faithful recreation of missing values is not the goal of imputation. In any case, several studies (e.g., Rässler, 2004) have shown promising results of multiple imputation when compared to other statistical matching methods.

Another popular imputation method is the Expectation-Maximization (EM) algorithm, which consists of two steps: expectation, which calculates the log-likelihood given some imputed values, and maximization, which maximizes the log-likelihood by adjusting the imputed values. EM is considered the best off-the-shelf method aside from multiple imputation, and has been extensively used in statistical matching applications (e.g., Kamakura and Wedel, 1997). However, Rässler (2002) showed that in the context of file concatenation, EM rarely converged to the global maximum. Further, it did not produce unique solutions, as the imputed values and model parameters were highly dependent on the algorithm's starting conditions, unless auxiliary data were present.

Finally, machine learning methods have been gradually making inroads into statistical matching, usually by embedding into existing frameworks. D'Orazio (2011) implemented tree-based machine learning algorithms – classification and regression tree (CART; Breiman et al., 1984), random forest (RF; Breiman, 2001), and recursive tree (Horton et al., 2006) – in a statistical matching application to find them capturing non-linearity in a synthetic dataset well. CART and RF have also been included in the popular R library for multiple imputation, *MICE*, where they can be used as univariate imputation methods (van Buuren, 2012). However, to date, machine learning methods are still largely absent from statistical matching applications (Putten and Kok, 2010). Transfer learning, by contrast, utilizes primarily machine learning methods, which we discuss in more detail in the next section.

No matter what framework or method for statistical matching is chosen, we should better understand the risks to validity, error propagation, and quality of inference in the fused data (Hand, 2018).

C.1.2 Overview of Machine Learning Methods Implemented in This Study

Technically not a machine-learning algorithm per se, *linear regression* provides a simple yet powerful and interpretable model of how inputs X affect outputs Y. The method may surpass more complicated, non-linear models in cases of small numbers of observations, sparse data, and low signal-to-noise ratio (Hastie et al., 2009). However, two particular challenges often arise in situations where "wide" datasets (having many variables or columns) are considered. First, by increasing the number of parameters (variable coefficients) in $f(\cdot)$, we overfit the model to the training dataset and sacrifice its transferability (the so-called *variance-bias tradeoff*, in which an overfit model reduces the bias involved in using a simpler model to reflect a more complex reality, but increases the variance between predicted and actual values when transferring the model to a new context). Second, the interpretability of such a model suffers because of the clutter created by copious parameters with associated marginal effects.

To overcome these challenges, *subset selection* and *shrinkage* methods are used in practice. Subset selection is a discrete approach in which variables are selected based on their performance in the model. The *best-subset selection* method searches the entire combinatorial space to pick the best performing specification. However, under current computational constraints, *best-subset selection* quickly becomes infeasible as the number of input variables increases. *Forward*- and *backward-stepwise selection* methods test each variable and at each stage include (exclude) the variable that most improves (least reduces) the fit until convergence at the given threshold is reached. Shrinkage methods offer a continuous solution to the problem of overspecification. Instead of the discrete choice of dropping or retaining a variable coefficient, they introduce an additive penalty term into the model, e.g., $\lambda \sum_{j=1}^{p} \beta_j^2$ in the case of *ridge regression* (where the β_j are parameters or variable coefficients, p is the number of parameters, and λ is the shrinkage operator), which is estimated simultaneously with the model and which prevents the large coefficient magnitudes that are common in the presence of multicollinearity (i.e., it shrinks the coefficient magnitudes toward zero). The penalty term for *Least Absolute Shrinkage and*

Selection Operator (*LASSO) regression*, $\lambda \sum_{j=1}^{p} |\beta_j|$, is similar to the one for ridge regression and also shrinks coefficient magnitudes; however, its non-linear nature allows *LASSO* to take the best of both discrete and continuous methods: by allowing some coefficients to shrink to zero (unlike ridge regression), it can effectively perform subset selection as well as shrinkage, which is essential for high-dimensionality problems (Hastie et al., 2009). Zou and Hastie (2005) proposed a convex combination of ridge and LASSO regression – the *elastic net*. Its penalty, $\lambda \sum_{j=1}^{p} \left(\alpha\beta_j^2 + (1-\alpha)|\beta_j|\right)$, is a generalization, which yields ridge or LASSO when $\alpha = 1$ or $\alpha = 0$, respectively.

If kNN attempts to create complex non-parametric boundaries between observations in hyperspace, multiple machine-learning methods, such as *logistic regression* and *linear (quadratic) discriminant analysis*, tackle this problem by imposing a functional relationship between X and Y. However, this separation could be one of infinitely many and may not be optimal, thus leading to misclassification of the new data points that map close to the class boundaries. *Support Vector Machines* (*SVMs*) are a method for classification and regression that solves this problem by finding optimal separating hyperplanes, that is, boundaries with the widest margins between classes. SVMs can handle inseparable problems and minimize the overlap of classes. Additionally, usage of kernel functions (e.g., *n*th-degree polynomials, radials, and neural networks) allows for creating non-linear boundaries in the original hyperspace.

Decision trees for classification and regression problems are a staple in the machine-learning field. The name refers to the way a model of this type is presented: it is a type of directed acyclic graph (DAG; Pelikan et al., 2001) with several nodes, each denoting a (usually binary) split based on the value of an explanatory variable. A tree starts with the first single split. Its child nodes may be iteratively split again, with the variables and split-points potentially differing by node. Tracing a branch down to a terminal node represents a set of conditions defining a group of observations that are predicted to have a certain value of the dependent variable, $\hat{y}$. Another way to illustrate the model is to partition the variable space into a set of volumes, each of which would correspond to a single (averaged) value of Y. These partitions are bounded by the split-planes that are equivalent to the binary split-nodes in the tree representation. The task of the model estimation algorithm, then, is to find independent variables X and their split-points that result in a more accurate prediction of Y.

Among the "tree-growing" methods used in practice, the *classification and regression tree* (*CART*; Breiman et al.,1984) and *C5.0* (including its earlier versions *ID3* and *C4.5*; Quinlan, 2014) are considered to be the "classic" algorithms. Varying in small details, they both use a greedy heuristic to do iterative splitting, and implement a tree-pruning cost function to avoid overspecification. Other variations include *recursive binary partitioning* (Horton et al., 2006), which employs non-parametric modeling, and globally-optimal trees obtained via *evolutionary learning* (Grubinger et al., 2014).

As summarized by Hastie et al. (2009), decision trees possess several desirable qualities: Decision trees can handle mixtures of continuous, ordered, and categorical independent variables naturally. They are insensitive to irrelevant inputs (e.g., case IDs, comment fields, etc.), and can treat item non-response as an explanatory variable. Scaling and other monotone transformations of the data do not affect the performance of tree models, nor does the presence of egregious outliers. Decision trees are computationally scalable and adequately interpretable. However, there are several issues common to the method:

- Trees are inherently unstable, noisy, very sensitive to the training data, and display high variance of the prediction.
- Trees are inherently unstable, noisy, very sensitive to the training data, and display high variance of the prediction.
- In regression problems, lack of smoothness could degrade performance of the learning function, if it is assumed to be smooth.
- Multiway (rather than the more common binary) splits are possible but they rapidly fragment and deplete data for the next level down.
- Additive structures are increasingly difficult to capture as the number of additive effects grows.

Ensemble methods in machine-learning try to alleviate one of the most severe shortcomings of the decision trees – overfitting to the training data. One approach involves a resampling method of bootstrap aggregation, or *bagging*. Bagging averages the prediction $\hat{Y}$ over multiple bootstrap samples, thereby reducing the variance of the fitted values and improving their accuracy at the price of losing the interpretable model structure. Moreover, bagging is beneficial only in cases when the unbagged model specification is not optimal, e.g., when parameters that cause overfitting are non-zero (Breiman, 1996). *Random forests* (Breiman, 2001) offer a substantial improvement over bagging procedures, by building multiple "de-correlated" decision trees. By randomly picking a subset of independent variables to grow each tree in the ensemble (hence, the name), the algorithm decreases the correlation between trees and, therefore, the variance of the averaged fitted values. The random forests approach is a quite popular, ready-to-use machine-learning method that requires very little tuning and produces robust results if the ratio of relevant to all variables in the dataset is not small.

Boosting, also an ensemble method, takes a different approach by assembling the "voting committee" from separate models. *Adaptive boosting*, or *AdaBoost* (Freund and Schapire, 1997), is the most widely used boosting algorithm. The algorithm relies on producing successive "weak" learners (decision trees), whose fitted outcomes could be just slightly better than a random guess. After each fitting iteration, observations in the training dataset are reweighted based on how well the model has performed: weights for observations that were predicted correctly by the previous learner are decreased, and weights for observations with unsuccessful predictions are increased. This procedure encourages each ensuing model to pay more attention to the difficult-to-predict cases. After models are estimated, their prediction results are averaged, while weighting the *results* from the more accurate models more heavily (in contrast to the *observation* weights mentioned above) to increase the influence of the better learners.

Extreme gradient boosting (*XGB;* Chen and Guestrin, 2016) is another popular ensemble method that incorporates a highly-scalable, sparsity-aware gradient tree algorithm. Its advantages include computational efficiencies through imbedded parallelization, caching, and approximation search; and an ability to handle sparse data natively; all of which bolstered usage of XGB in applications in digital advertising, insurance, and particle physics.

It is widely agreed in the machine-learning community that boosting is one of the best general machine-learning treatments available off-the-shelf. In many cases, boosting, which has decision trees serving as the basic algorithm, helps to overcome the main drawback of decision trees – inaccuracy due to their large variance – at the price of lower computational speed and some reduction in interpretability.

Another large group of methods includes *artificial neural networks* (*ANNs*), which are known for being universal function approximators (Hornik, 1991). Although ANNs consist of very simple elementary units, interconnected architecture with fully, partially, and recursively connected "hidden" layers could render it very complex (the infamous "black-box"). Recent advancements in network design and hardware have allowed even larger and more complex models (with millions of neurons) in the pursuit of prediction accuracy (Goodfellow et al., 2016). As such, rectified linear unit and other novel activation functions were introduced to prevent gradient decay while training deep networks. At the same time, graphics processing units (GPUs) provided massive parallelization yielding multiple fold training time decrease compared to conventional central processing units (CPUs). This enabled the sharp rise of ANN applications in computer vision, speech recognition, natural language processing, and artificial intelligence. This list, however, shows that ANNs are extremely successful in domains with homogenous input such as images, sounds, and spoken and written language.

Apparently, there is no silver bullet when it comes to selecting the best algorithm for the task: its performance is contingent on various factors, such as nature, content, and representation of the explanatory and dependent variables, and the strength of associations between them. In this study, we test all aforementioned machine-learning algorithms. For comparative performance tables, refer to Table C.8 and Table C.9.

C.2 Transfer Learning Results for the Classification Problem

Selected cross-validation results of the classification tasks are presented in Table C.1 (the full results can be found in Table C.9 – Cross-validation results for the classification problem). The CV-aided search for the best learner is similar to the one presented for the regression problem, with the exception of a few details. Specifically, instead of MSE, the misclassification error (MCE) is used as the generalization error. K-nearest neighbors (kNN) and forward stepwise multinomial logit (MNL) are added to the pool of the learning functions. Assigning the mean value and CART learners are replaced by assigning the median value and C5.0, respectively. For the forward stepwise linear regression learner, predicted values are rounded to the nearest eligible integer that represents a response category. Finally, artificial neural networks (ANNs) and extreme gradient boosting (XGB) are added to the pool of the tested learning functions.

Across 15 learning functions, both kNN and XGB achieve the minimum MCE for 14 out of 39 dependent variables each, while LASSO regression (MNL kernel) delivers minimum MCE only for 11 dependent variables. However, after averaging the MCE results over all dependent variables, LASSO regression has a slight edge over kNN and XGB – 0.567, 0.569, and 0.570, respectively (indicating correct classification of 433, 431, or 430 observations out of 1,000). Overall, the results demonstrate that even the best-performing learners do a rather inadequate job in predicting attitudes given the available inputs: on average, LASSO regression predictions are only ΔMCE=0.034 more accurate (i.e., with only 34 more correct per 1,000 observations) than assigning the median.

Table C.1 – Cross-validation results for the classification problem

Variable	**Best learner**	**Lowest MCE**	**Median assignment MCE**	**ΔMCE (median assignment vs. best learner)**
A1a_goodcommute	Median/ LASSO/ SVM/ AdaBoost/ kNN	0.433	0.433	0.000
A1b_jobmoney	kNN	0.623	0.628	–0.005
A1c_closestore	AdaBoost/ kNN	0.536	0.537	–0.001
A1d_prefdrive	kNN/ ANN	0.673	0.751	–0.078
A1e_boring	Median/ kNN/ XGB/ ANN	0.577	0.577	0.000
A1f_deadline	XGB	0.555	0.557	–0.002
A1g_yards	XGB	0.590	0.619	–0.029
A1h_newtech	kNN	0.589	0.590	–0.001
A1i_traffic	LASSO	0.588	0.602	–0.014
A1j_transit	LASSO	0.560	0.583	–0.023
A1k_trendset	XGB	0.629	0.663	–0.034
A1l_dayoff	kNN/ XGB/ ANN	0.546	0.547	–0.001
A1m_grocery	LASSO	0.487	0.546	–0.059
A1n_timetowork	XGB	0.572	0.575	–0.003
A1o_eproducts	Recursive tree	0.620	0.630	–0.010
A1p_travelwaste	XGB	0.557	0.569	–0.012
A1q_stresscommute	LASSO	0.509	0.511	–0.002
A1r_goodjob	Median/ LASSO/ SVM/ kNN	0.407	0.407	0.000
A1s_walkbike	LASSO	0.627	0.786	–0.159
A1t_closetransit	kNN	0.614	0.764	–0.150
A1u_liketravel	Recursive tree/ AdaBoost/ kNN/ XGB/ ANN	0.492	0.493	–0.001
A1v_useminute	XGB	0.638	0.720	–0.082
A1w_techproblems	MNL/ LASSO	0.644	0.709	–0.065
A1x_destination	AdaBoost/ kNN	0.450	0.451	–0.001
A1y_payquicktrip	SVM	0.645	0.683	–0.038
A1z_transitovercar	LASSO	0.602	0.766	–0.164
A1aa_noisyshops	AdaBoost	0.593	0.749	–0.156
A1ab_hurry	Recursive tree	0.650	0.737	–0.087
A1ac_carjustmove	LASSO/ XGB/ ANN	0.498	0.500	–0.002
A1ad_wanttravel	SVM	0.639	0.823	–0.184
A1ae_likeinternet	kNN	0.465	0.467	–0.002
A1af_driving	XGB	0.543	0.546	–0.003
A1ag_busy	ANN	0.532	0.535	–0.003
A1ah_impressivecar	XGB	0.636	0.645	–0.009
A1ai_fewtrips	XGB	0.487	0.494	–0.007
A1aj_welcomecommute	Recursive tree/ kNN	0.594	0.628	–0.034
A1ak_wbovercar	LASSO	0.503	0.553	–0.050
A1al_neverbehind	kNN	0.630	0.782	–0.152
A1am_goodlife	XGB	0.414	0.416	–0.002

As in the regression problem, the generalization errors vary across the board for the attitudinal variables. For only nine variables out of 39 is the MCE lower than 0.5, meaning that prediction success is achieved for more than 50% of the test observations. These variables load on the factors *travel is wasted time* (3 variables), *satisfaction* (2), *commute benefit* (1), *pro-active transportation* (1), and *pro-technology* (1), in addition to the statement "*I prefer to organize my errands so that I make as few trips as possible*", which is not included in the factor analysis. However, in all but one case the prediction rate is attained mainly because of the variable distributions (demonstrating extreme peakedness) rather than the performance of a sophisticated learning function: the MCE deviation from the median assignment learner is 0 (i.e., the median is the best or one of the best predicting functions) or very close to 0. Only for the statement, "*I like the idea of living in a neighborhood where I can walk to the grocery store*", is the MCE 0.487 for the LASSO regression, which constitutes a 0.059 improvement over assigning the median. When the distribution of $\mathcal{Y}$ is not very peaked, conditional learning functions could noticeably improve the generalization errors. For example, the SVM learner gains 0.184 of ΔMCE over the median for the statement "*I sometimes travel more than I have to, because I want to*", indicating that the explanatory variables improved the prediction of this variable.

C.3 Supporting Tables

Table C.2 – Attitudinal statements in the general opinions section of the MSNCC (N=2,849)

Variable name	**Statement**	**Median value**[a]
A1a_goodcommute	My commute is generally pleasant.	Agree
A1b_jobmoney	The main benefit of my job is that it gives me the money to pay for the things I *really* enjoy doing.	Agree
A1c_closestore	When I need to buy something, I usually prefer to get it at the closest store possible.	Agree
A1d_prefdrive	I'd rather drive than travel by any other means.	Neutral
A1e_boring	The act of traveling is boring.	Disagree
A1f_deadline	I feel more productive when I am under pressure to complete work by a deadline.	Agree
A1g_yards	I like the idea of living somewhere with large yards and lots of space between homes.	Agree
A1h_newtech	I like to track the development of new technology.	Agree
A1i_traffic	Getting stuck in traffic doesn't bother me much.	Disagree
A1j_transit	I like the idea of transit as a means of travel for me.	Agree
A1k_trendset	I often introduce new trends to my friends.	Neutral
A1l_dayoff	Occasionally, I'd be willing to give up a day's pay to get a day off work.	Agree
A1m_grocery	I like the idea of living in a neighborhood where I can walk to the grocery store.	Agree
A1n_timetowork	I do my best work when I have more than enough time to complete it.	Agree
A1o_eproducts	I like to be among the first to own new electronic products.	Disagree
A1p_travelwaste	Time spent traveling is generally wasted time.	Disagree
A1q_stresscommute	My commute is stressful.	Disagree
A1r_goodjob	I am generally satisfied with my job.	Agree
A1s_walkbike	I prefer to walk or bike rather than drive whenever possible.	Neutral
A1t_closetransit	I prefer to live close to transit, even if it means I'll have a smaller home and more people living nearby.	Neutral
A1u_liketravel	I generally enjoy the act of traveling itself.	Agree
A1v_useminute	I feel like I need to make the most of every single minute.	Neutral
A1w_techproblems	Technology brings at least as many problems as solutions.	Neutral
A1x_destination	The only good thing about traveling is arriving at your destination.	Disagree
A1y_payquicktrip	I would pay money to reduce the time I spend traveling.	Neutral
A1z_transitovercar	I prefer to take transit rather than drive whenever possible.	Neutral

Table C.2 (Continued)

Variable name	**Statement**	**Median value**
A1aa_noisyshops	Mixing different types of businesses (e.g., shops, restaurants, offices) with the homes in my neighborhood causes (or would cause) too much traffic or noise.	Neutral
A1ab_hurry	I'm often in a hurry to be somewhere else.	Neutral
A1ac_carjustmove	To me, a car is mostly just a way to get from place to place.	Agree
A1ad_wanttravel	I sometimes travel more than I *have* to, because I *want* to.	Neutral
A1ae_likeinternet	The internet makes life more interesting.	Agree
A1af_driving	I like the idea of driving as a means of travel for me.	Agree
A1ag_busy	I'm too busy to do many things I'd like to do.	Agree
A1ah_impressivecar	I (would) like to own a car that impresses other people.	Disagree
A1ai_fewtrips	I prefer to organize my errands so that I make as few trips as possible.	Agree
A1aj_welcomecommute	My commute serves as a welcome transition between home and work.	Agree
A1ak_wbovercar	I like the idea of walking (or biking) as a means of transportation.	Agree
A1al_neverbehind	I never get very far behind on things I'm trying to get done.	Neutral
A1am_goodlife	I am generally satisfied with my life.	Agree

[a] Reporting scale has five levels: “Strongly disagree”, “Disagree”, “Neutral”, “Agree”, and “Strongly Agree”.

Table C.3 – General attitudinal latent constructs (factors)

Constructs[a]	**Statements**	**Pattern matrix loadings**[b]
Pro-transit [AVT9_protransit][c]	I prefer to take transit rather than drive whenever possible.	0.739
	I'd rather drive than travel by any other means.	–0.588
	I like the idea of driving as a means of travel for me.	–0.536
	I like the idea of transit as a means of travel for me.	0.510
Travel is wasted time [AVT9_nec_oftravel]	I generally enjoy the act of traveling itself.	–0.774
	The act of traveling is boring.	0.710
	Time spent traveling is generally wasted time.	0.592
	The only good thing about traveling is arriving at your destination.	0.567
	I sometimes travel more than I have to, because I want to.	–0.389
	To me, a car is mostly just a way to get from place to place.	0.308
Pro-technology [AVT9_protech]	I like to be among the first to own new electronic products.	0.755
	I like to track the development of technology.	0.747
	I often introduce new trends to my friends.	0.577
	The internet makes life more interesting.	0.343
	Technology brings at least as many problems as solutions.	–0.305
Commute benefit [AVT9_comm_ben]	My commute is generally pleasant.	0.773
	My commute is stressful.	–0.769
	My commute serves as a welcome transition between home and work.	0.372
Time pressure – reality [AVT9_timepres_real]	I'm often in a hurry to be somewhere else.	0.674
	I'm too busy to do many things I'd like to do.	0.476
	I feel like I need to make the most of every single minute.	0.433
Time pressure – preference [AVT9_timepres_pref]	I do my best work when I have more than enough time to complete it.	–0.709
	I feel more productive when I am under pressure to complete work by a deadline.	0.532
Pro-active transportation [AVT9_pro_activetrans]	I like the idea of walking (or biking) as a means of transportation.	0.895
	I prefer to walk or bike rather than drive whenever possible.	0.767
	I like the idea of living in a neighborhood where I can walk to the grocery store.	0.420
Satisfaction [AVT9_satisfaction]	I am generally satisfied with my life.	0.806
	I am generally satisfied with my job.	0.550
Pro-density [AVT9_prodensity]	I like the idea of living somewhere with large yards and lots of space between homes.	–0.635
	I prefer to live close to transit, even if it means I'll have a smaller home and more people living nearby.	0.625
	Mixing different types of businesses (e.g., shops, restaurants, offices) with the homes in my neighborhood causes (or would cause) too much traffic or noise.	–0.549

[a] Principal axis factor extraction with oblimin rotation was used.
[b] Represents the degree of association between the statement and the construct. Only loadings greater than 0.3 in magnitude are reported.
[c] Variable name in the input/output datasets.

Table C.4 – Socio-economic variables common to the MSNCC and NHTS datasets

NHTS variable name	Variable content (following the NHTS)	Variable description (following the NHTS)
HH_HISP	Hispanic	Binary
HH_RACE	Race	Categorical: 1=White, 2=Black, 3=Asian, 4=Native American, 5=Native Hawaiian, 6=Multi ethnic
DRVRCNT	Number of drivers in HH	Count
HHFAMINC	Derived total annual HH income	Ordinal: 1 to 16=$0k to $80k w/ $5k increments, 17=$80-100k, 18= >$100k
HHSIZE	Count of HH members (HHMs)	Count
HHVEHCNT	Count of HH vehicles	Count
NUMADLT	Count of adult HHMs at least 18 years old	Count
WRKCOUNT	Number of workers in HH	Count
LIF_CYC	Life cycle classification for the HH	Categorical: 1 to 10=combination of adults and children of various age categories.
BORNINUS	Respondent was born in U.S.	Binary
CONDNIGH	Medical condition results in limiting driving to daytime	Binary
CONDPUB	Medical condition results in using bus/subway less frequently	Binary
DRIVER	Driver status of respondent	Binary
EDUC	Highest grade completed	Ordinal: 1= <HS, 2=HS, 3=Some college, 4=Bachelor's, 5=Graduate degree
GCDWORK	Great circle distance (miles) between home and work	Continuous
OCCAT	Job category	Categorical: 1=Sales/service, 2=Clerical/admin, 3=Manufacturing, 4=Profess./managerial, 97=Other
R_AGE	Respondent age (years)	Continuous
R_SEX	Respondent gender	Binary: 1=Male, 2=Female
SELF_EMP	Self-employed	Binary
TIMETOWK	Minutes to go from home to work last week	Continuous
TRAVDAY	Travel day – day of week	Categorical: 1=Sunday,…, 7=Saturday
WKFTPT	Work full or part-time	Binary: 1=Full-time, 2=Part-time
WORKER	Respondent worker status	Binary
WRKTRANS	Transportation mode to work last week	Categorical: 1=Car, 2=Van, 3=SUV, 4=Pickup truck, 5=Other truck, 6=RV, 7=Motorcycle, 8=Light EV, 9=Local bus, 10=Commuter bus, 11=School bus, 12=Charter bus, 13=Intercity bus, 14=Shuttle bus, 15=Amtrak, 16=Commuter train, 17=Subway/elevated, 18=Streetcar, 19=Taxi, 20=Ferry, 21=Airplane, 22=Bicycle, 23=Walk, 24=Spec transit, 97=Other
DISTTOWK	One-way distance to workplace (miles)	Continuous
TDAYDATE	Date of travel day (YYYYMM)	Date

Table C.5 – Comparison of selected variable distributions in the source (NHTS) and target (MSNCC) domains

Variable	Mean		Test stat.[a]	p-value
	Source	Target		
HH_HISP	0.08	0.08	0.014	0.90
HH_RACE: White	0.66	0.86	694.175	0.00
HH_RACE: Black	0.04	0.05	6.268	0.01
HH_RACE: Asian	0.15	0.03	1072.830	0.00
HHFAMINC: $0-25k	0.08	0.07	0.065	0.80
HHFAMINC: $25-50k	0.14	0.20	45.504	0.00
HHFAMINC: $50-75k	0.19	0.20	1.141	0.29
HHFAMINC: $75-100k	0.19	0.19	0.001	0.97
HHFAMINC: >$100k	0.38	0.29	70.913	0.00
DRIVER	0.96	0.98	47.174	0.00
EDUC: less than HS degree	0.00	0.04	79.752	0.00
EDUC: HS degree	0.03	0.23	530.803	0.00
EDUC: less than BS/BA degree	0.24	0.29	29.938	0.00
EDUC: BS/BA degree	0.31	0.24	62.672	0.00
EDUC: graduate degree	0.42	0.18	853.179	0.00
OCCAT: service	0.06	0.25	433.764	0.00
OCCAT: clerical	0.15	0.12	9.878	0.00
OCCAT: manufacture	0.02	0.13	282.541	0.00
OCCAT: professional	0.64	0.44	404.836	0.00
R_SEX: female	0.61	0.49	127.523	0.00
WORKER	0.94	1.00	1821.530	0.00
WRKTRANS: car	0.51	0.94	6204.507	0.00
WRKTRANS: motorcycle	0.01	0.01	0.076	0.78
WRKTRANS: local bus	0.06	0.01	435.505	0.00
WRKTRANS: express bus	0.07	0.00	1969.312	0.00
WRKTRANS: heavy rail	0.08	0.01	1560.817	0.00
WRKTRANS: light rail	0.16	0.00	7757.984	0.00
WRKTRANS: bicycle	0.10	0.01	2485.862	0.00
WRKTRANS: walk	0.02	0.02	0.160	0.69

Table C.5 (continued)

Variable	Mean		Test stat.[a]	p-value
	Source	Target		
DRVRCNT	2.20	2.23	0.097	0.00
HHSIZE	2.70	2.93	0.086	0.00
HHVEHCNT	2.04	2.58	0.214	0.00
NUMADLT	1.97	2.23	0.138	0.00
WRKCOUNT	2.19	1.80	0.163	0.00
R_AGE	43.87	47.57	0.139	0.00
TIMETOWK	44.66	24.24	0.343	0.00
DISTTOWK	21.16	14.29	0.150	0.00

[a] Kolmogorov-Smirnoff and chi-square statistics are used for continuous and categorical variables, respectively.

Table C.6 – Descriptive statistics of the observed continuous attitudes for the MSNCC (N=2,352)

Variable	Mean	SD	Median	Min	Max	Skew	Kurtosis
Pro-transit	–0.01	1.00	–0.05	–3.06	2.76	0.03	–0.46
Travel is wasted time	0.01	1.00	–0.15	–2.39	3.62	0.46	0.03
Pro-technology	–0.01	1.01	–0.04	–2.85	2.95	0.01	–0.26
Commute benefit	0.00	1.00	0.23	–3.60	2.00	–0.76	0.47
Time pressure – reality	0.01	1.00	0.01	–3.06	3.94	0.03	–0.34
Time pressure – preference	0.02	0.99	–0.01	–2.85	3.02	0.03	–0.32
Pro-active transportation	0.00	1.00	0.15	–2.94	1.80	–0.45	–0.38
Satisfaction	0.01	1.00	0.09	–4.12	1.96	–0.89	1.27
Pro-density	0.00	1.01	–0.07	–3.02	2.77	0.16	–0.30

Table C.7 – Descriptive statistics of the predicted continuous attitudes for the MSNCC (N=2,352)

Variable	**Mean**	**SD**	**Median**	**Min**	**Max**	**Skew**	**Kurtosis**
Pro-transit	–0.01	0.46	–0.04	–1.12	1.21	0.15	–1.11
Travel is wasted time	0.01	0.10	0.01	–0.39	0.27	–0.33	–0.11
Pro-technology	–0.01	0.25	–0.02	–0.63	1.42	0.38	0.22
Commute benefit	0.00	0.34	0.00	–1.14	1.15	0.05	–0.37
Time pressure – reality	0.01	0.11	0.01	–0.38	0.55	0.06	–0.11
Time pressure – preference	0.02	0.21	0.04	–1.05	0.48	–0.28	–0.14
Pro-active transportation	0.00	0.45	–0.09	–1.05	1.53	0.96	0.53
Satisfaction	0.01	0.20	0.02	–1.06	1.17	–0.47	1.36
Pro-density	0.00	0.48	–0.05	–1.32	1.81	0.42	0.16

Table C.8 – Cross-validation results for the regression problem

Variable	**MSE**											**Best learner**	**ΔMSE (mean assignment vs. best learner)**
	RHD	**Assigning the mean**	**Forward stepwise linear regression**	**CART**	**Evolutionary regression tree**	**Recursive tree**	**Bagging**	**Random forest**	**LASSO**	**SVM**	**AdaBoost**		
Pro-transit	2.021	0.993	0.895	0.829	0.815	0.816	0.826	0.823	0.757	0.801	0.771	LASSO	–0.236
Travel is wasted time	1.952	1.001	1.146	1.003	1.000	0.992	1.071	1.078	0.985	1.005	1.012	LASSO	–0.016
Pro-technology	2.035	1.017	1.110	0.971	0.965	0.962	1.022	1.026	0.951	0.968	0.971	LASSO	–0.066
Commute benefit	2.080	1.008	1.279	0.932	0.930	0.904	0.959	0.963	0.898	0.961	0.919	LASSO	–0.110
Time pressure – reality	2.003	1.009	1.169	1.002	0.994	1.002	1.075	1.083	1.003	0.995	1.022	Evolutionary regression tree	–0.015
Time pressure – preference	1.903	0.994	1.112	0.963	0.968	0.955	1.004	1.005	0.936	0.953	0.954	LASSO	–0.058
Pro-active transportation	2.009	1.009	0.923	0.848	0.847	0.854	0.863	0.866	0.789	0.842	0.811	LASSO	–0.220
Satisfaction	1.904	1.004	1.156	0.995	0.991	0.994	1.045	1.046	0.976	0.993	0.996	LASSO	–0.028
Pro-density	1.970	1.005	0.848	0.847	0.828	0.829	0.831	0.837	0.748	0.762	0.761	LASSO	–0.257

Table C.9 – Cross-validation results for the classification problem

Variable	MCE							
	RHD	**Assigning the median**	**Forward stepwise linear regression**	**C5.0**	**Evolutionary classification tree**	**Recursive tree**	**Bagging**	**Random forest**
A1a_goodcommute	0.615	0.433	0.526	0.508	0.437	0.434	0.496	0.506
A1b_jobmoney	0.744	0.628	0.658	0.676	0.641	0.629	0.663	0.666
A1c_closestore	0.703	0.537	0.638	0.627	0.547	0.537	0.602	0.601
A1d_prefdrive	0.775	0.751	0.690	0.712	0.679	0.686	0.700	0.715
A1e_boring	0.705	0.577	0.647	0.667	0.608	0.582	0.646	0.648
A1f_deadline	0.703	0.557	0.691	0.655	0.576	0.563	0.617	0.630
A1g_yards	0.725	0.619	0.648	0.663	0.637	0.619	0.638	0.650
A1h_newtech	0.717	0.590	0.632	0.642	0.604	0.596	0.614	0.636
A1i_traffic	0.686	0.602	0.629	0.641	0.618	0.603	0.643	0.652
A1j_transit	0.709	0.583	0.591	0.612	0.601	0.586	0.598	0.614
A1k_trendset	0.722	0.663	0.648	0.685	0.676	0.656	0.664	0.663
A1l_dayoff	0.712	0.547	0.710	0.612	0.562	0.549	0.602	0.604
A1m_grocery	0.672	0.546	0.519	0.554	0.521	0.511	0.550	0.553
A1n_timetowork	0.688	0.575	0.631	0.660	0.590	0.575	0.630	0.631
A1o_eproducts	0.739	0.630	0.672	0.709	0.643	0.620	0.693	0.691
A1p_travelwaste	0.703	0.569	0.656	0.645	0.573	0.570	0.620	0.629
A1q_stresscommute	0.679	0.511	0.585	0.600	0.516	0.514	0.564	0.574
A1r_goodjob	0.585	0.407	0.493	0.483	0.412	0.408	0.475	0.479
A1s_walkbike	0.766	0.786	0.680	0.674	0.679	0.660	0.663	0.683
A1t_closetransit	0.775	0.764	0.655	0.672	0.647	0.638	0.653	0.656
A1u_liketravel	0.645	0.493	0.601	0.596	0.514	0.492	0.537	0.538

Table C.9 (continued)

Variable	MCE							
	RHD	Assigning the median	Forward stepwise linear regression	C5.0	Evolutionary classification tree	Recursive tree	Bagging	Random forest
A1v_useminute	0.729	0.720	0.696	0.691	0.686	0.644	0.684	0.683
A1w_techproblems	0.728	0.709	0.687	0.663	0.678	0.649	0.665	0.667
A1x_destination	0.626	0.451	0.583	0.528	0.458	0.454	0.507	0.516
A1y_payquicktrip	0.711	0.683	0.658	0.687	0.675	0.668	0.682	0.666
A1z_transitovercar	0.762	0.766	0.673	0.678	0.642	0.609	0.650	0.678
A1aa_noisyshops	0.736	0.749	0.674	0.685	0.627	0.618	0.647	0.651
A1ab_hurry	0.707	0.737	0.703	0.704	0.682	0.650	0.692	0.695
A1ac_carjustmove	0.674	0.500	0.663	0.594	0.509	0.500	0.572	0.572
A1ad_wanttravel	0.744	0.823	0.769	0.688	0.653	0.653	0.685	0.688
A1ae_likeinternet	0.618	0.467	0.485	0.533	0.484	0.467	0.520	0.524
A1af_driving	0.676	0.546	0.608	0.598	0.564	0.554	0.575	0.586
A1ag_busy	0.692	0.535	0.656	0.618	0.538	0.537	0.613	0.610
A1ah_impressivecar	0.740	0.645	0.665	0.691	0.664	0.648	0.698	0.696
A1ai_fewtrips	0.563	0.494	0.525	0.533	0.535	0.496	0.543	0.550
A1aj_welcomecommute	0.711	0.628	0.673	0.671	0.626	0.594	0.653	0.648
A1ak_wbovercar	0.705	0.553	0.581	0.596	0.517	0.513	0.560	0.592
A1al_neverbehind	0.723	0.782	0.735	0.678	0.663	0.662	0.667	0.666
A1am_goodlife	0.578	0.416	0.479	0.489	0.426	0.416	0.469	0.472

Table C.9 (continued)

Variable	MCE							Best learner	ΔMCE (median assignment vs. best learner)
	MNL	LASSO	SVM	AdaBoost	kNN	XGB	ANN		
A1a_goodcommute	0.487	0.433	0.433	0.433	0.433	0.434	0.434	Median/ LASSO/ SVM/ AdaBoost/ kNN	0.000
A1b_jobmoney	0.682	0.624	0.639	0.627	0.623	0.627	0.627	kNN	–0.005
A1c_closestore	0.594	0.537	0.545	0.536	0.536	0.537	0.538	AdaBoost/ kNN	–0.001
A1d_prefdrive	0.699	0.680	0.692	0.674	0.673	0.675	0.673	kNN/ ANN	–0.078
A1e_boring	0.639	0.583	0.588	0.579	0.577	0.577	0.577	Median/ kNN/ XGB/ ANN	0.000
A1f_deadline	0.593	0.560	0.560	0.562	0.556	0.555	0.556	XGB	–0.002
A1g_yards	0.642	0.609	0.612	0.602	0.597	0.590	0.614	XGB	–0.029
A1h_newtech	0.617	0.590	0.599	0.594	0.589	0.590	0.590	kNN	–0.001
A1i_traffic	0.611	0.588	0.604	0.617	0.595	0.592	0.598	LASSO	–0.014
A1j_transit	0.611	0.560	0.574	0.581	0.575	0.577	0.571	LASSO	–0.023
A1k_trendset	0.678	0.642	0.659	0.649	0.630	0.629	0.643	XGB	–0.034
A1l_dayoff	0.601	0.547	0.552	0.547	0.546	0.546	0.546	kNN/ XGB/ ANN	–0.001
A1m_grocery	0.524	0.487	0.505	0.501	0.495	0.500	0.514	LASSO	–0.059
A1n_timetowork	0.627	0.575	0.587	0.585	0.576	0.572	0.573	XGB	–0.003
A1o_eproducts	0.658	0.629	0.635	0.625	0.629	0.626	0.631	Recursive tree	–0.010
A1p_travelwaste	0.619	0.566	0.578	0.564	0.562	0.557	0.566	XGB	–0.012
A1q_stresscommute	0.563	0.509	0.517	0.515	0.510	0.510	0.511	LASSO	–0.002
A1r_goodjob	0.449	0.407	0.407	0.408	0.407	0.409	0.409	Median/ LASSO/ SVM/ kNN	0.000
A1s_walkbike	0.649	0.627	0.678	0.649	0.658	0.677	0.674	LASSO	–0.159
A1t_closetransit	0.652	0.615	0.621	0.620	0.614	0.625	0.645	kNN	–0.150
A1u_liketravel	0.528	0.493	0.498	0.492	0.492	0.492	0.492	Recursive tree/ AdaBoost/ kNN/ XGB/ ANN	–0.001

Table C.9 (continued)

Variable	MCE							Best learner	ΔMCE (median assignment vs. best learner)
	MNL	LASSO	SVM	AdaBoost	kNN	XGB	ANN		
A1v_useminute	0.706	0.645	0.654	0.647	0.648	0.638	0.642	XGB	–0.082
A1w_techproblems	0.664	0.644	0.670	0.656	0.651	0.655	0.654	MNL/ LASSO	–0.065
A1x_destination	0.511	0.452	0.458	0.450	0.450	0.453	0.452	AdaBoost/ kNN	–0.001
A1y_payquicktrip	0.665	0.647	0.645	0.657	0.654	0.653	0.662	SVM	–0.038
A1z_transitovercar	0.656	0.602	0.654	0.607	0.652	0.673	0.628	LASSO	–0.164
A1aa_noisyshops	0.651	0.602	0.603	0.593	0.601	0.598	0.605	AdaBoost	–0.156
A1ab_hurry	0.682	0.660	0.677	0.670	0.664	0.670	0.670	Recursive tree	–0.087
A1ac_carjustmove	0.541	0.498	0.507	0.501	0.499	0.498	0.498	LASSO/ XGB/ ANN	–0.002
A1ad_wanttravel	0.669	0.660	0.639	0.651	0.640	0.642	0.650	SVM	–0.184
A1ae_likeinternet	0.489	0.467	0.471	0.469	0.465	0.466	0.466	kNN	–0.002
A1af_driving	0.586	0.550	0.557	0.556	0.546	0.543	0.546	XGB	–0.003
A1ag_busy	0.580	0.535	0.543	0.535	0.535	0.535	0.532	ANN	–0.003
A1ah_impressivecar	0.648	0.639	0.652	0.637	0.641	0.636	0.638	XGB	–0.009
A1ai_fewtrips	0.522	0.496	0.510	0.524	0.502	0.487	0.489	XGB	–0.007
A1aj_welcomecommute	0.639	0.597	0.608	0.601	0.594	0.596	0.597	Recursive tree/ kNN	–0.034
A1ak_wbovercar	0.536	0.503	0.547	0.513	0.538	0.553	0.517	LASSO	–0.050
A1al_neverbehind	0.667	0.632	0.650	0.640	0.630	0.634	0.648	kNN	–0.152
A1am_goodlife	0.448	0.417	0.415	0.415	0.417	0.414	0.415	XGB	–0.002

REFERENCES

Abrantes, Pedro A. L. and Mark R. Wardman (2011) Meta-analysis of UK values of travel time: an update. *Transportation Research A* **45**, 1-17.

Ahfock, Daniel, Saumyadipta Pyne, Sharon X. Lee, and Geoffrey J. McLachlan (2016) Partial identification in the statistical matching problem. *Computational Statistics & Data Analysis* **104**, 79-90.

American Association of State Highway and Transportation Officials (AASHTO) (2015) *Commuting in America 2013. The National Report on Commuting Patterns and Trends.* Washington, DC. Available at http://traveltrends.transportation.org/Pages/default.aspx, accessed April 3, 2015.

Anderson, James M., Kalra Nidhi, Karlyn D. Stanley, Paul Sorensen, Constantine Samaras, and Oluwatobi A. Oluwatola (2014) *Autonomous Vehicle Technology: A Guide for Policymakers*. Santa Monica, California: Rand Corporation. Available at http://www.rand.org/pubs/research_reports/RR443-1.html, accessed July 20, 2014.

Andridge, Rebecca R. and Roderick J.A. Little (2010) A review of hot deck imputation for survey non-response. *International Statistical Review* **78(1)**, 40-64.

Antonelli, Joseph, Corwin Zigler, and Francesca Dominici (2017) Guided Bayesian imputation to adjust for confounding when combining heterogeneous data sources in comparative effectiveness research. *Biostatistics* **18(3)**, 553-568.

Babbie, Earl (2010) *The Practice of Social Research*, 12th edition. Belmont, CA: Wadsworth Publishing Company.

Baker, Ken, Paul Harris and John O'Brien (1989) Data fusion: An appraisal and experimental evaluation. *Journal of the Market Research Society* **31(2)**, 153-212.

Bastian, Anne, Maria Börjesson, Jonas Eliasson (2016) Explaining “peak car” with economic variables. *Transportation Research A* **88**, 236–250.

Bastian, Anne, Maria Börjesson, Jonas Eliasson (2017) Response to Wadud and Baierl: “Explaining ‘peak car’with economic variables: An observation”. *Transportation Research A* **95**, 386-389.

Becker, Gary S. (1965) A theory of the allocation of time. *Economic Journal* **75**, 493-517.

Becker, Jan C. and Andreas Simon (2000) Sensor and navigation data fusion for an autonomous vehicle. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2000*, 156-161. IEEE.

Bellman, Richard E. (1961) *Adaptive Control Processes*. Princeton University Press.

Ben-Akiva, Moshe E. and Steven R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, Massachusets: MIT Press.

Berliner, Rosaria M., Aliaksandr Malokin, Giovanni Circella, and Patricia L. Mokhtarian (2015) Travel-based multitasking: modeling the propensity to conduct activities while commuting. Presented at the 94th Transportation Research Board Meeting, Washington, DC.

Bluedorn, Allen. C., Thomas J. Kalliath, Michael J. Strube and Gregg D. Martin (1999) Polychronicity and the Inventory of Polychronic Values (IPV). *Journal of Managerial Psychology* **14(3/4)**, 205-230.

Bluedorn, Allen C., and Gwen Martin (2008) The time frames of entrepreneurs. *Journal of Business Venturing* **23(1)**, 1-20.

Blumenberg, Evelyn, Brian D. Taylor, Michael Smart, Kelcie Ralph, Madeline Wander and Stephen Brumbaugh (2012) *What's Youth Got to Do With It? Exploring the Travel Behaviors of Teens and Young Adults.* Los Angeles: UCTC-FR-2012-14.

Breiman, Leo (1996) Bagging predictors. *Machine Learning* **24**, 123-140.

Breiman, Leo (2001) Random forests. *Machine Learning* **45**, 5-32.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984) *Classification and Regression Trees*. Monterey, CA: Wadsworth & Brooks.

Brownstone, David and Kenneth A. Small (2005) Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research A* **39**, 279-293.

Cameron, A. Colin and Pravin K. Trivedi (2005) *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cao, Xinyu, Patricia L. Mokhtarian, and Susan L. Handy (2007) Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. *Environment and Planning A* **39(4)**, 830-847.

Cao, Xinyu, Patricia L. Mokhtarian, and Susan L. Handy (2009) Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews* **29(3)**, 359-395.

Castanedo, Federico (2013) A review of data fusion techniques. *The Scientific World Journal* **2013**, 1-19.

Chen, Aiyou, Art B. Owen, and Minghui Shi (2015) Data enriched linear regression. *Electronic Journal of Statistics* **9(1)**, 1078-1112.

Chen, Tianqi and Carlos Guestrin (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. ACM.

Chesley, Noelle (2014) Information and communication technology use, work intensification and employee strain and distress. *Work, Employment & Society,* doi: 10.1177/0950017013500112. Available at http://wes.sagepub.com/content/early/2014/03/10/0950017013500112.abstract, accessed July 30, 2014.

Circella, Giovanni, Patricia L. Mokhtarian and Laura K. Poff (2012) A conceptual typology of multitasking behavior and polychronicity preferences. *electronic International Journal of Time Use Research* **9(1)**, 59-107.

Connolly, Dara, Brian Caulfield, and Margaret O'Mahony (2009) Rail passengers' preferences for on-board Wi-fi internet access. Transportation Research Board Paper #09-2618, Washington, DC, January.

Daly, Andrew and Stephane Hess (2019) VTT or VTTS: a note on terminology for value of travel time work. *Transportation*, available at https://doi.org/10.1007/s11116-018-9966-4, accessed on March 21, 2019.

Delbosc, Alexa and Graham Currie (2013) Causes of youth licensing decline: A synthesis of evidence. *Transport Reviews* **33(3)**, 271-290.

DeLoach, Stephen B. and Thomas K. Tiemann (2012) Not driving alone? American commuting in the twenty-first century. *Transportation* **39**, 521-537.

DeSerpa, Allan C. (1973) Microeconomic theory and the valuation of travel time: Some clarification. *Regional and Urban Economics* **2**, 401-410.

Dong, Zhi, Patricia L. Mokhtarian, Giovanni Circella and James R. Allison (2015) The estimation of changes in rail ridership through an onboard survey: did free Wi-Fi make a difference to Amtrak's Capitol Corridor service? *Transportation* **42**(1), 123-142.

D'Orazio, Marcello, Marco Di Zio, and Mauro Scanu (2006) *Statistical Matching: Theory and Practice*. Chichester, England: John Wiley & Sons.

Du, Ke-Lin and M. N. S. Swamy (2013) *Neural Networks and Statistical Learning*. London: Springer.

El Faouzi, Nour-Eddin, Henry Leung, and Ajeesh Kurian (2011) Data fusion in intelligent transportation systems: Progress and challenges – A survey. *Information Fusion* **12(1)**, 4-10.

Etezady, Ali, Patricia L. Mokhtarian, Giovanni Circella, and Aliaksandr Malokin (2019) Not all minutes are created equal: How does the impact of travel time on commute mode choice differ by demographics and propensity to multitask? Paper #19-03326 presented at the 98th Annual Meeting of the Transportation Research Board, Washington, DC, January.

Ettema, Dick and Laura Verschuren (2007). Multitasking and value of travel time savings. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2010, pp. 19-25.

Ettema, Dick, Margareta Friman, Tommy Gärling, Lars E. Olsson, and Satoshi Fujii (2012) How in-vehicle activities affect work commuters' satisfaction with public transport. *Journal of Transport Geography* **24**, 215-222.

FHWA (Federal Highway Administration) (2016) Travel monitoring: traffic volume trends. Available at https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm, accessed on July 21 2016.

Fisseler, Jens and Imre Fehér (2009) Data fusion with probabilistic conditional logic. *Logic Journal of IGPL* **18(4)**, 488-507.

Fosdick, Bailey K., Maria DeYoreo, and Jerome P. Reiter (2016) Categorical data fusion using auxiliary information. *The Annals of Applied Statistics* **10(4)**, 1907-1929.

Frei, Charlotte, Hani S. Mahmassani, and Andreas Frei (2015) Making time count: Traveler activity engagement on urban transit. *Transportation Research A* **76**, 58-70.

Freund, Yoav and Robert E. Schapire (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* **55(1)**, 119-139.

Garikapati, Venu M., Ram M. Pendyala, Eric A. Morris, Patricia L. Mokhtarian and Noreen McDonald (2016) Activity patterns, time use and travel of millennials: a generation in transition? *Transport Reviews* **36(5)**, 558-584.

Giuliano, Genevieve (2005) Low income, public transit, and mobility. *Transportation Research Record* **1927**, 63-70.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016) *Deep Learning*. Cambridge: MIT press.

Gravina, Raffaele, Parastoo Alinia, Hassan Ghasemzadeh, and Giancarlo Fortino. (2017) Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion* **35**, 68-80.

Gripsrud, Mattias and Randi Hjorthol (2012) Working on the train: from ‘dead time’ to productive and vital time. *Transportation* **39**, 941-956.

Grubinger, Thomas, Achim Zeileis, and Karl-Peter Pfeiffer (2014) evtree: Evolutionary learning of globally optimal classification and regression trees in R. *Journal of Statistical Software* **61(1)**, 1-29.

Gunn, Hugh (2001) Spatial and temporal transferability of relationships between travel demand, trip cost and travel time. *Transportation Research E* **37**, 163-189

Guo, Zhan, Alexandra Derian and Jinhua Zhao (2015) Smart devices and travel time use by bus passengers in Vancouver, Canada. *International Journal of Sustainable Transportation* **9(5)**, 335-347.

Guttman, Joel M. (1975) Implicit assumption and choices among estimates of the value of time. *Transportation Research Record* **534**, 63-68

Hall, Dave L. and James Llinas (1997) Introduction to multisensor data fusion. *Proceedings of IEEE* **85(1)**, 6-23.

Hand, David J. (2018) Statistical challenges of administrative and transaction data. *Journal of the Royal Statistical Society, Series A* **181(3)**, 1-24.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman (2009) *The Elements of Statistical Learning; Data Mining, Inference, and Prediction. Second Edition*. New York: Springer.

Hauser, John R. (1978) Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach. *Operations Research* **26(3)**, 406-421.

Hensher, David A. and Mark Bradley (1993) Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters* **4(2)**, 139-151.

Hensher, David A. and Lester W. Johnson (1981) *Applied Discrete-Choice Modelling*. London: Croon-Helm.

Hensher, David A. and Baojin Wang (2016) Productivity foregone and leisure time corrections of the value of business travel time savings for land passenger transport in Australia. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice* **25(2)**, 15-29.

Hopkins, Debbie and Janet Stephenson (2015) *Generation Y Mobilities: Full Report*. Centre for Sustainability (CSAFE), University of Otago.

Hornik, Kurt (1991) Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4(2)**, 251-257.

Horton, Torsten, Kurt Hornik, and Achim Zeileis (2006) Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics* **15(3)**, 651-674.

Hüttenrauch, Bettina (2016) *Targeting Using Augmented Data in Database Marketing: Decision Factors for Evaluating External Sources*. Wiesbaden, Germany: Springer.

Kamakura, Wagner A. and Michel Wedel (1997) Statistical data fusion for cross-tabulation. *Journal of Marketing Research* **34(4)**, 485-498.

Kenyon, Susan and Glenn Lyons (2007) Introducing multitasking to the study of travel and ICT: Examining its extent and assessing its potential importance. *Transportation Research A* **41(2)**, 161-175.

Keogh, Eamonn and Abdullah Mueen (2010) Curse of dimensionality. In *Encyclopedia of Machine Learning*, ed. Claude Sammut and Geoffrey I. Webb, 257-258. Boston: Springer.

Kerr, Alistair, Alexia Lennon and Barry Watson (2010) The call of the road: factors predicting students' car travelling intentions and behavior. *Transportation* **37(1)**, 1-13.

Keseru, Imre, and Cathy Macharis (2017) Travel-based multitasking: review of the empirical evidence. *Transport Reviews*, http://dx.doi.org/10.1080/01441647.2017.1317048.

Khaleghi, Bahador, Alaa Khamis, Fakhreddine O. Karray, and Saiedeh N. Razavi (2013) Multisensor data fusion: A review of the state-of-the-art. *Information Fusion* **14(1)**, 28-44.

König, Cornelius. J. and Mary J. Waller. (2010) Time for reflection: A critical examination of polychronicity. *Human Performance* **23(2)**, 173-190.

Koppelman, Frank S. and Chandra Bhat (2006) *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models*. Available at http://caee.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf, accessed on Aug. 29, 2018.

Kouwenhoven, Marco and Gerard de Jong (2018) Value of travel time as a function of comfort. *Journal of Choice Modelling* **28**, 97-107.

Kuhnimhof, Tobias, Jimmy Armoogum, Ralph Buehler, Joyce Dargay, Jon Denstali and Toshiyuki Yamamoto (2012) Men shape a downward trend in car use amongst young adults – Evidence from six industrialized countries. *Transport Reviews* **32(6)**, 761-779.

Lagani, Vincenzo, Sofia Triantafillou, Gordon Ball, Jesper Tegner, and Ioannis Tsamardinos (2016) Probabilistic computational causal discovery for systems biology. In *Uncertainty in Biology*, ed. Liesbet Geris and David Gomez-Cabrero, 33-73. Cham, Switzerland: Springer.

Lindquist, Jay D. and Carol Kaufman-Scarborough (2007) The Polychronic-Monochronic Tendency Model: PMTS scale development and validation. *Time & Society* **16**, 253-285.

Lyons, Glenn and John Urry (2005) Travel time in the information age. *Transportation Research A* **39(2-3**), 257-276.

Malokin, Aliaksandr, Giovanni Circella and Patricia L. Mokhtarian (2015) How do activities conducted while commuting influence mode choice? Testing public transportation advantage and autonomous vehicle scenarios. Presented at the 94th Transportation Research Board Meeting, Washington, DC.

Malokin, Aliaksandr, Giovanni Circella and Patricia L. Mokhtarian (2017a) Do multitasking millennials value travel time differently? A revealed preference study of Northern California commuters. Presented at the 96th Transportation Research Board Meeting, Washington, DC. Currently under journal review; available from the authors.

Malokin, Aliaksandr, Patricia L. Mokhtarian, and Giovanni Circella (2017b) *An Investigation of Methods for Imputing Attitudes from One Sample to Another*. School of Civil and Environmental Engineering, Georgia Institute of Technology, Research Report. Available at http://hdl.handle.net/1853/58418, accessed July 12, 2017.

Malokin, Aliaksandr, Giovanni Circella and Patricia L. Mokhtarian (2018a) Do multitasking millennials value travel time differently? A revealed preference study of Northern California commuters. Available from the authors.

Malokin, Aliaksandr, Patricia L. Mokhtarian and Giovanni Circella (2018b) An investigation of methods to enrich National Household Travel Survey data with attitudinal variables. Available from the authors.

Malokin, Aliaksandr, Giovanni Circella and Patricia L. Mokhtarian (2019) How do activities conducted while commuting influence mode choice? Using revealed preference models to inform public transportation advantage and autonomous vehicle scenarios. *Transportation Research Part A* **124**, 82-114.

McDonald, Noreen C. (2015) Are millennials really the "go-nowhere" generation? *Journal of the American Planning Association*, **81(2)**, 90–103.

Mishra, Gouri S., Patricia L. Mokhtarian and Keith F. Widaman (2015) An empirical investigation of attitudes toward waiting on the part of Northern California commuters. *Travel Behaviour and Society,* **2(2)**, 78-87.

Mokhtarian, Patricia L. (2016) Discrete choice models' ρ^2: A reintroduction to an old friend. *Journal of Choice Modelling* **21**, 60-65.

Mokhtarian, Patricia L. (2018) The times they are a-changin': what do the expanding uses of travel time portend for policy, planning, and life? Thomas B. Deen Distinguished Lecture of the Annual Meeting of the Transportation Research Board. *Transportation Research Record*, https://journals.sagepub.com/doi/10.1177/0361198118798602.

Mokhtarian, Patricia L., Francis Papon, Matthieu Goulard, and Marco Diana (2015) What makes travel pleasant and/or tiring? An investigation based on the French National Travel Survey. *Transportation* **42**, 1103-1128, doi: 10.1007/s11116-014-9557-y.

Mokhtarian, Patricia L. and Ilan Salomon (2001) How derived is the demand for travel? Some conceptual and measurement considerations. *Transportation Research Part A* **35**, 695-719.

Neufeld, Amanda J. and Patricia L. Mokhtarian (2012) *A Survey of Multitasking by Northern California Commuters: Description of the Data Collection Process*. Institute of Transportation Studies, Research Report UCD-ITS-RR-12-32. Available at http://www.its.ucdavis.edu/?page_id=10063&pub_id=1802, accessed April 1, 2017.

Newcombe, Howard B., James M. Kennedy, S. J. Axford and Allison P. James (1959) Automatic linkage of vital records. *Science* **130(3381)**, 954-959.

Newgard, Craig D., Jerris R. Hedges, Melanie Arthur, and Richard J. Mullins (2004) Advanced statistics: The propensity score – a method for estimating treatment effect in observational research. *Academic Emergency Medicine* **11(9)**, 953-961.

NHTSA (2013) *U.S. Department of Transportation releases policy on automated vehicle development*. Available at http://www.nhtsa.gov/About+NHTSA/Press+Releases/U.S.+Department+of+Transportation+Releases+Policy+on+Automated+Vehicle+Development, accessed July 6, 2016.

Ohmori, Nobuaki and Noboru Harata (2008) How different are activities while commuting by train? A case in Tokyo. *Tijdschrift voor Economische en Sociale Geografie* **99(5)**, 547-561.

Okner, Benjamin (1974) Data matching and merging: an overview. *Annals of Economic and Social Measurement* **3(2)**, 347-352.

Pan, Sinno Jialin and Qiang Yang (2010) A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22(10)**, 1345-1359.

Pew Research Center (2014) *Millennials in adulthood. Detached form institutions, networked with friends.* Pew Research Center, Washington, D.C. Available at http://www.pewsocialtrends.org/files/2014/03/2014-03-07_generations-report-version-for-web.pdf, accessed on September 19, 2015.

Rasouli, Soora and Harry Timmermans (2014) Judgments of travel experiences, activity envelopes, trip features and multi-tasking: A panel effects regression model specification. *Transportation Research Part A: Policy and Practice* **63**, 67-75.

Rässler, Susanne (2002) *Statistical Matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York: Springer Science & Business Media.

Rässler, Susanne (2004) Data fusion: identification problems, validity, and multiple imputation. *Austrian Journal of Statistics* **33(1-2)**, 153-171.

Rhee, Kyoung-Ah, Joon-Ki Kim, Back-Jin Lee, Sungyop Kim, Young-Ihn Lee (2013) Analysis of effects of activities while traveling on travelers' sentiment. *Transportation Research Record* **2381**, 27-34.

Rosenbloom, Sandra and Elizabeth Burns (1994) Why working women drive alone: Implications for travel reduction programs. *Transportation Research Record* **1459**, 39-45.

Rubin, Donald B. (1986) Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics* **4(1)**, 87-94.

Schifeling, Tracy, Jerome P. Reiter, and Maria DeYoreo (2016) Data Fusion for Correcting Measurement Errors. arXiv preprint arXiv:1610.00147. Available at https://arxiv.org/pdf/1610.00147.pdf, accesed on April 9, 2018.

Shannon, Tya, Billie Giles-Corti, Terri Pikora, Max Bulsara, Trevor Shilton and Fiona Bull (2006) Active commuting in a university setting: Assessing commuting habits and potential for modal change. *Transport Policy* **13(3)**, 240-253.

Shaw, F. Atiyya, Aliaksandr Malokin, Patricia L. Mokhtarian, and Giovanni Circella (2018) It's not all fun and games: an investigation of the reported benefits and disadvantages of conducting activities while commuting. Under peer review; available from the authors.

Shires, Jeremy D. and Gerard C. de Jong (2009) An international meta-analysis of values of travel time savings. *Evaluation and Program Planning* **32**, 315-325.

Singleton, Patrick A. (2018) How useful is travel-based multitasking? Evidence from Portland, Oregon commuters. *Transportation Research Record.* https://doi.dox.org/10.1177/0361198118776151.

Sivak, Michael and Brandon Schoettle (2011) *Recent Changes in the Age Composition of Drivers in 15 Countries*. Michigan: UMTRI-2011-43.

Susilo, Y. O, G. Lyons, J. Jain and S. Atkins (2012) Rail passengers' time use and utility assessment. *Transportation Research Record* **2323(1)**, 99-109.

Swait, Joffre and Jordan Louviere (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* **30(3)**, 305-314.

Talvittie, Antti (1972) Comparison of probabilistic modal choice models: estimation methods and system inputs. *Highway Research Record* **392**, 111-120.

Tang, Jia, Feng Zhen, Jason Cao, and Patricia L. Mokhtarian (2018) How do passengers use travel time? A case study of Shanghai-Nanjing High Speed Rail. *Transportation* **45(2)**, 451-477.

Taniguchi, Kyle Ken (2012) *Work Trips on Public Transportation: An Analysis of Trends, Select Markets, and Users Using the National Household Travel Survey Series*. MS Thesis, Civil and Environmental Engineering, University of South Florida. Available at http://scholarcommons.usf.edu/etd/4411/, accessed October 4, 2013.

Train, Kenneth (2009) *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press. Available at https://eml.berkeley.edu/books/choice2.html, accessed November 16, 2017.

Trottenberg, Polly. Memorandum to secretarial officers, modal administrators (2011) *Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis*, September 28. Available at http://www.dot.gov/sites/dot.gov/files/docs/Value%20of%20Travel%20Time%20Memorandum.pdf, accessed April 3, 2015.

Tsamardinos, Ioannis, Sofia Triantafillou, and Vincenzo Lagani (2012) Towards integrative causal analysis of heterogeneous data sets and studies. *Journal of Machine Learning Research* **13(Apr)**, 1097-1157.

Tully, Claus (2011) Mobilisierung des Mobilen – Trends in der Jugendmobilität. *Der Nahverkehr* **29(7-8)**, 12–15.

U.S. Census Bureau (2011) *Decennial Census 2010*. Available at https://www2.census.gov/, accessed on April 1, 2017.

U.S. Census Bureau (2014) *American Community Survey 2013, 5-year estimates*. Available at https://www2.census.gov/, accessed on April 1, 2017.

U.S. Department of Transportation, Federal Highway Administration (2009) *National Household Travel Survey*. Available at http://nhts.ornl.gov, accessed on April 1, 2017.

U.S. Environmental Protection Agency (2014) *Smart Location Database, 2013*. Available at https://www.epa.gov/smartgrowth/smart-location-mapping, accessed on April 1, 2017.

van Buuren, Stef (2012) *Flexible Imputation of Missing Data*. Boca Raton, Florida: CRC press.

van der Putten, Peter and Joost N. Kok (2010) Using data fusion to enrich customer databases with survey data for database marketing. In *Marketing Intelligent Systems Using Soft Computing*, ed. Jorge Casillas and Francisco J. Martínez-López, 113-130. Berlin: Springer.

van der Putten, Peter, Joost N. Kok, and Amar Gupta (2002) Data fusion through statistical matching. MIT Sloan Working Paper No. 4342-02. Available at http://liacs.leidenuniv.nl/~puttenpwhvander/library/2002fusionsloan.pdf, accessed June 10, 2017.

van der Waerden, P., A. Kemperman, H. Timmermans and R. van Hulle (2010) The influence of facilities for multitasking on individual's travel decisions in the context of work trips. Transportation Research Board Paper #10-2459, Washington, DC, January.

Varghese, Varun and Arnab Jana (2018) Impact of ICT on multitasking during travel and the value of travel time savings: Empirical evidences from Mumbai, India. *Travel Behaviour and Society* **12**, 11-22.

VCloudNews. Every day big data statistics – 2.5 quintillion bytes of data created daily. Available at http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/, accessed on May 31, 2017.

Wadud, Zia and Baierl, Martin (2017) Explaining "peak car" with economic variables: A comment. *Transportation Research Part A* **95**, 381-385.

Wagner, Jason, Trey Baker, Ginger Goodin, and John Maddox (2014) *Automated Vehicles: Policy Implications Scoping Study*. Texas A&M Transportation Institute, Texas A&M University, Research Report SWUTC/14/600451-00029-1. Available at http://d2dtl5nnlpfr0r.cloudfront.net/swutc.tamu.edu/publications/technicalreports/600451-00029-1.pdf, accessed July 20, 2014.

Wallis, Claudia (2010) The impacts of media multitasking on children's learning and development – Report from a research seminar, The Joan Ganz Cooney Center and Stanford University, available at http://www.joanganzcooneycenter.org/Reports-22.html, accessed July 23, 2014.

Waltz, Edward (1986) Data fusion for C3I: A tutorial. In *Command, Control, Communications Intelligence (C3I) Handbook*, prepared by the editors of Defense Electronics, 217-226. Palo Alto, CA: EW Communications.

Wardman, Mark and Glenn Lyons (2016) The digital revolution and worthwhile use of travel time: implications for appraisal and forecasting. *Transportation* **43**, 507-530.

Watts, Laura (2008) The art and craft of train travel. *Social and Cultural Geography* **9(6)** (September), 711-726.

Watts, Laura and John Urry (2008) Moving methods, travelling times. *Environment & Planning D* **26**, 860-874.

White, Franklin E. (1991) *Data Fusion Lexicon*. San Diego, U.S.: Joint Directors of Laboratories, Technical Panel for C3, Data Fusion Sub-Panel, Naval Ocean Systems Center. Available at http://www.dtic.mil/dtic/tr/fulltext/u2/a529661.pdf, accessed on May 3, 2018.

Winkler, William E. (1999) *The State of Record Linkage and Current Research Problems*. Technical Report, Statistical Research Division, U.S. Census Bureau.

Wu, Ge, Toshiyuki Yamamoto, and Ryuichi Kitamura (1999) Vehicle ownership model that incorporates the causal structure underlying attitudes toward vehicle ownership. *Transportation Research Record* **1676**, 61-67.

Zamparini, Luca and Aura Reggiani (2007) Meta-analysis and the value of travel time savings: a transatlantic perspective in passenger transport. *Networks and Spatial Economics* **7(4)**, 377-396.

Zhang, Junyi and Harry Timmermans (2010) Scobit-based panel analysis of multitasking behavior of public transport users. *Transportation Research Record* **2157**, 46-53.

Zheng, Yu (2015) Methodologies for cross-domain data fusion: An overview. *IEEE Transactions on Big Data* **1(1)**, 16-34.

Zheng, Zuduo, Simon Washington, Paul Hyland, Keith Sloan, and Yulin Liu (2016) Preference heterogeneity in mode choice based on a nationwide survey with a focus on urban rail. *Transportation Research Part A: Policy and Practice* **91**, 178-194.

Zhou, Jiangping (2012) Sustainable commute in a car-dominant city: Factors affecting alternative mode choices among university students. *Transportation Research A* **46(7)**, 1013-1029.

Zorn, Christopher (2005) A solution to separation in binary response models. *Political Analysis* **13(2)**, 157-170.